Circuit Theory and Design

STEWART

WILEY

— Circuit Theory and Design —

STEWART

50

Circuit Theory and Design

Circuit Theory and Design

JOHN L. STEWART
Associate Professor of Electrical Engineering
California Institute of Technology

5/29/57

JOHN WILEY & SONS, INC., NEW YORK
Chapman & Hall, Limited, London

Library of Congress Catalog Card Number: 56-8696

Printed in the United States of America

Preface

The majority of devices of concern to the electrical and electronics engineer can, at least to a first order, be described as linear. Virtually all systems have characteristics that vary in some manner with frequency. Unquestionably, the most powerful technique for understanding and controlling linear, frequency-dependent systems is that afforded by the poles and zeros of network functions. In addition to the general conceptual and analytical tool which they provide, their manipulation also yields an exceptionally effective tool for network design. In electric and mechanical circuits both with and without energy sources, the pole-zero approach tends to make analysis and design one and the same; so powerful is the technique in providing understanding (analysis) that only a slight change in viewpoint leads to design (synthesis).

Although pole-zero methods have been employed by high-level systems designers and network synthesis people for several years, the technique has not become nearly as widespread as is justified. In fact, its greatest value appears to lie outside the realm of formal network synthesis, where it has apparently resided all this time. Because a descriptive and relatively thorough text has not been available, the student, as well as his teacher, has remained partially unaware of pole-zero methods. This book has been written to help fill the void. It includes considerable information and design procedures of practical significance and discusses a wide variety of specific devices. As background, the student should have had calculus and if possible a course in elementary differential equations, although the latter is not necessary. In addition, the student should have taken the usual sophomore and junior courses on circuit theory and preferably an introductory course on vacuum tubes. To facilitate the use of this book as a text, a number of problems have been provided for each chapter. To a large extent, they relate to practical system design. In addition, the problems present many useful formulas and concepts.

It has been my express intent to write this book so that it might prove effective in undergraduate courses without compromising its value in

graduate study. Although the introduction of poles and zeros in other than a cursory fashion into the undergraduate curriculum is somewhat revolutionary at this time, I nevertheless believe it is inevitable. The problem involved in carrying out my intent was therefore reduced to the manner of organizing the material. In order to point out how I feel the book can best be used, a brief résumé of its content is warranted.

Chapter 1 is a relatively thorough and generalized review of steady-state circuit analysis. The student may have been exposed to much of this material, although some of it, such as determinant manipulation, may be new. I have adopted the less common convention of associating the negative sign with all mutual terms in node and loop equations, which has many advantages. The student learns the form of a general system of equations from relatively simple examples, which always show the negative sign. It is cumbersome and confusing to the student to define a fiction such as coimpedance in order to write plus in front of off-diagonal terms. In order to apply determinental methods, the negative sign must ultimately be set down anyway.

Poles and zeros are first introduced in Chapter 2, employing a strict phasor approach, and the behavior of functions is emphasized in terms of their poles and zeros. In Chapter 3, the relationships between networks and poles and zeros are discussed. In addition, a fairly extensive introduction to transient calculations is given, with a Heaviside approach. The concepts developed in Chapters 2 and 3 are freely used in the balance of the book.

Chapter 4 deals with *R-L*, *R-C*, and *L-C* networks and their canonical forms, and also dwells at considerable length on the extremely important concepts of impedance and frequency normalization and frequency transformations. By assuming a knowledge of normalization and transformation throughout the balance of the book, it has been possible to make the later parts more compact and less cluttered than is normally possible.

Chapter 5 is a fairly detailed study of maximally flat, linear-phase, and other functions, maintained fairly independent of specific networks. This chapter is, in essence, a treatise on the approximation problem. Some of the functions developed in Chapter 5 are often presented at later points in the book without additional discussion.

Chapter 6 introduces some of the more important topics from modern network synthesis, for example, a simplified form of the Darlington procedure. It is not intended as a substitute for a formal graduate course in modern network theory, although it does provide a good starting point.

Chapters 7 and 8 are devoted to the classical theory of image matching in somewhat more mathematical detail than is customary in undergraduate texts. A reliance on a knowledge of frequency transformations and normalization has greatly reduced the amount of space necessary to cover this material adequately.

Chapter 9 is a study of vacuum-tube linear equivalent circuits and some special phenomena associated with tubes. Although the level of this chapter is somewhat lower than the rest of the book, its inclusion is felt to be justified for reference, completeness, and to facilitate the use of the text in the undergraduate curriculum. Chapters 7 and 8 serve a similar purpose in some respects.

Chapters 10 and 11 go into the design of low-pass and band-pass amplifiers in some detail, with the aim of developing relatively simple methods for designing amplifier systems which are today considered rather sophisticated. Poles and zeros are employed extensively in these chapters. In addition, the step-function response of low-pass amplifiers is studied in more than a cursory fashion.

The final three chapters are devoted to feedback devices. Chapter 12 is concerned with feedback amplifiers, and emphasizes the precision design of the closed-loop transfer function. Chapter 13 takes up linear oscillators, generalized to a considerable extent in relation to the associated circuitry. Chapter 14 is concerned with the functions of interest in servomechanism theory and design. As in Chapter 12, precision design is emphasized. This chapter, along with the rest of the book, provides a much more thorough background for the serious study of linear servomechanisms than is usual.

As stated, I have tried to organize the subject matter so that the book can be used at either the graduate or the undergraduate level. How I feel the book can best be used at the graduate level will first be discussed.

The usual graduate program has three distinct courses related to circuits: a one-year study of formal network theory, a one-year consideration of electronic circuit design, and a one-semester course on transient analysis. In this book I have tried to combine practical linear circuit design with much of the material contained in formal network theory. No attempt has been made to integrate a detailed study of transients into the presentation, although the material in Chapter 3 is adequate for most practical problems. Rather, it is felt best to leave this to an independent course. In addition, most of the formality and rigor of network theory has been supplanted with more readily understandable pole-zero and phasor concepts.

This book can be covered in a one-year course, especially if Chap-

ters 1, 4, 7, 8, or 9 are treated as review material. In addition to this book, the first-year graduate program can be made complete along the lines of circuit and network theory by adding the usual formal course on Fourier and Laplace theory, a one-semester course on nonlinear and other special aspects of electronic circuits, and, finally, a one-semester treatment of formal network theory, which is better appreciated by a student who has the first half of this book as background than by one with an earlier formal network theory course. A student of network and circuit theory has always been hampered by the lack of an easily grasped pictorial representation; poles and zeros interpreted as phasors can furnish this lack.

The possible adoption of this book in the undergraduate program has been my fondest hope. I am convinced that it can be used effectively, providing suitable instructors are available to teach it. Of course, it is always hard to modify an undergraduate curriculum, which is a difficulty that must be accepted. However, the book has been written with this difficulty in mind. The first eight chapters, with the probable exception of Chapter 6, can be introduced most easily through a modification of the customary course on classical filter theory, which normally has about a year of circuit analysis as a prerequisite. Chapters 1, 4, 7, and 8 include all the material usually taught in such courses, which means that only Chapters 2, 3, and 5 represent new material. If the first half of the book serves as a text, the introduction of the second half into the curriculum is no great problem; Chapters 9 through 14 can be covered in one semester.

Unfortunately, the new graduate does not appear to know much circuit design, and may even have trouble with relatively simple analysis problems. One reason for this is the great bulk of concepts forced upon him without a detailed study of any one, which is a recent malady arising from the tremendous expansion of frontiers. Teaching too many concepts is questionable philosophy; it is much easier to learn concepts after graduation than to learn techniques of analysis and design involving mathematical and even simple algebraic manipulations. This is of particular concern in the field of linear circuit theory and design, which constitutes what is probably the engineer's most important special knowledge. His lack of training in circuit theory is compounded because the classical approach does not leave him with a thorough and deep-seated understanding of networks. The pictorial approach given by poles and zeros is particularly suited to providing the needed understanding. However, practice and repeated application to design problems are necessary if the student is to retain his knowledge; in other words, a cursory introduction to the subject is not adequate.

The undergraduate and many graduates find it difficult to grasp circuit theory thoroughly if all proofs and manipulations are purely mathematical; they need some picture in their minds. For this reason, the formal justification of poles and zeros with Laplace, Fourier, or comparable mathematical disciplines has been ignored. In addition, analogs such as electrolytic tanks and rubber membranes have been ignored; I do not feel they give any more understanding than that obtainable from a simple phasor approach. Wherever possible, the phase and amplitude characteristics of a circuit have been implied with a simple pole-zero diagram in hopes that the student will begin to think and visualize in these terms.

An effort has been made to show that the variable p has many interpretations, all of which can be considered simultaneously; frequency variable, complex variable, derivative operator, and Heaviside (or Laplace) operator. The result of this emphasis on the multitudinous nature of p has been a loose functional notation for voltages and currents. It would be unfortunate to restrict the interpretation of a voltage by writing a symbol such as $E(j\omega)$, which is normally reserved for the steady state.

I will no doubt be accused of using symbols in a few places that do not conform to certain standards set down a number of years ago. The first is the way I define plate and grid voltages. I have not flaunted standard notation; rather, I have avoided it. In all but the simplest vacuum-tube equivalent circuits, all node voltages must be taken with respect to the reference node if utter confusion is not to result. Thus, I use e_v for grid-to-reference-node (ground) voltage and e_a for plate-to-ground voltage. I would prefer to use the more familiar symbols e_g and e_p; however, these standard symbols refer voltage to the cathode. It is only in the special case that the cathode also happens to be the reference node. Thus the symbols e_g and e_p are, unfortunately, relatively useless. In many diagrams, I use rather arbitrary symbols for e_v and e_a, such as e_0, e_3, and e_L, which avoids a cumbersome double-subscript notation. All too often, a strict adherence to standards is equivalent to a strait jacket.

In addition, I have not used the customary symbols for voltage and current generators. It is indeed unfortunate if the student cannot tell the two apart without an extra label. The convention showing a circle with an arrow alongside is particularly annoying (same symbol for both current and voltage); it breaks down completely when applied to, say, a source of force or velocity. The symbols I have adopted are unique, include polarity, do not require an extra lable, clutter up a diagram to the least possible extent, and have an obvious interpretation; a current

source is a circle with an arrow inside, and a voltage source is a circle with plus and minus signs inside.

Little in the way of specific numerical designs will be found here except in the problems. The intent has been to present general methods from which any specific design is merely a straightforward application of the more general theory. This philosophy has been extended to the figures as well; where specific tube-characteristic and other curves might have been employed, generalized and qualitative sketches have been given instead. In any event, the inclusion of numerical examples should be the responsibility of the teacher and problem assignments rather than the text.

A separate bibliography has been provided with comments on the references in a chapter-by-chapter fashion. This obviates the need for footnotes and permits a more detailed account to be given of the scope of the various references as related to the material here.

Analyses with poles and zeros interpreted as phasors are little documented. Much of the work here is a result of personal effort and as such may show ignorance of what may have been done by others at an earlier date. The origin of the pole-zero method is difficult to ascertain; it may have been employed as a conceptual tool by such men as Gauss and Maxwell. Records of relatively early applications to circuit theory do not seem to be available except in association with normal modes and in analogy to potential functions. Neither of these interpretations has been emphasized here.

The origin of this book stemmed from conversations between Dr. D. O. Pederson, University of California, and myself in early 1953. From its very inception, Dr. H. H. Skilling, Chairman of the electrical engineering department at Stanford University, has provided considerable encouragement and enthusiasm, without which this book probably would not have been written. I am grateful to Professors J. M. Pettit and D. F. Tuttle, Jr., who introduced me to pole-zero concepts while I was a student at Stanford University in 1948–1949. Faculty members at the University of Michigan have been helpful in their comments on and enthusiasm for the work; in particular, I am grateful to Professor S. S. Attwood, Chairman of the electrical engineering department, who made typing facilities available for earlier versions of the text. Most of all, I am indebted to my wife Rita for her encouragement and patience through many long evenings of work.

JOHN L. STEWART

Pasadena, California
July, 1956

Contents

I

Circuit Analysis

Network analysis is of basic importance to the study of lumped networks with or without vacuum tubes. If some driving function is applied to a known circuit, the output of the circuit is calculable by means of an analysis. Although it may be tedious and time-consuming, analysis of lumped and linear networks never involves more than a straightforward application of mathematical rules. The present chapter is devoted to the basic principles of analysis. It is assumed that the material in this chapter will not be entirely new to the reader; rather, it has been included to provide review, to set forth certain definitions, and to formalize systems of network equations.

Primarily, this book is devoted to the synthesis and design of networks, both with and without vacuum tubes. Synthesis can be broadly grouped into two spheres of effort, both of which are related to the construction of some network such that the output from the network will be a prescribed function of the known input. Function synthesis involves the derivation of a relatively abstract mathematical function that is realizable with a network and that is desirable from some point of view. In accomplishing function synthesis, various mathematical techniques can be employed and perfectly good functions can be derived by mathematicians who may not be too conversant with analysis and circuits in general. We shall have a considerable amount to say about this topic later.

The other kind of synthesis is network synthesis, which is devoted to the attainment of a physical circuit that realizes some given function. This second kind of synthesis is so closely related to analysis that it becomes difficult to draw a sharp dividing line between the two. For example, a designer may possess a repertoire of circuit knowledge that he has gathered through experience. By adjusting the parameters of these circuits, he may be able to design a rather broad group of circuit functions. Contrasted to this, formal network synthesis involves the application of a set of rules to a given mathematical function, the end product being a network. Like function synthesis, formal network

synthesis does not require appreciable practical experience in circuit design. Unfortunately, formal procedures all too often yield impractical networks. Formal network synthesis has not been developed to the point where vacuum tubes can be treated in a straightforward manner, although it may be quite useful for many parts of vacuum-tube networks. We shall study the more prosaic method of network synthesis in greatest detail but shall occasionally have a word to say regarding formal procedures.

Before treating specific topics, it is well to specify the types of networks to be studied in this book and the restrictions that will be imposed. The systems to be studied here, with two or three exceptions, will be linear with parameters that do not change with time. The equations with which we shall deal will be *linear integrodifferential equations with constant coefficients*. This means that all variables have constant multipliers; the variables as well as their derivatives and integrals never appear to powers other than the first; and products of variables do not appear. Admittedly, this is but a first-order approximation to actual systems but it is often a good one. The deviation from linearity in circuit elements is usually hard to detect. Vacuum tubes are not linear but can be approximated as such, with the accuracy improving as the amplitudes of the signals handled by the tubes decrease. Frequently, the assumption of linearity is not accurate but can be used to obtain approximate solutions to problems without resorting to a nonlinear analysis which may be impractical or impossible to apply.

In addition, the studies in this book will be restricted to networks having lumped elements not subject to radiation effects. This means that inductors, resistors, and capacitors can be localized and treated as if they existed at points rather than over spatially sizable regions. Neglecting radiation is tantamount to restricting our studies to circuit configurations that are small in size compared to the wavelengths of the voltages and currents circulating in the networks such that a change at one point in the network is instantaneously felt at every other point.

1.1 Network elements and mutual inductance

There exist four different network parameters, two of which are quite similar. They are resistance R, capacitance C, inductance L, and mutual inductance M. R, L, and C are defined by the voltage across their

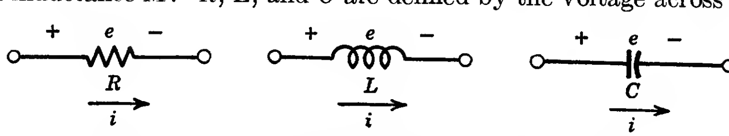


Fig. 1.1. Current-voltage relations in circuit elements.

terminals $e(t)$ and the current that flows through them $i(t)$ (see Fig. 1.1) according to

$$\begin{aligned} e(t) &= Ri(t) \\ e(t) &= L\frac{di(t)}{dt} \\ e(t) &= S\int i(t)\,dt = S\int_0^t i(t)\,dt + e_0 \end{aligned} \tag{1.1}$$

where $S = 1/C$ is called the elastance and the constant e_0 is the voltage across the capacitor at the reference time $t = 0$.

The inverse relations are

$$\begin{aligned} i(t) &= Ge(t) \\ i(t) &= C\frac{de(t)}{dt} \\ i(t) &= \Gamma\int e(t)\,dt = \Gamma\int_0^t e(t)\,dt + i_0 \end{aligned} \tag{1.2}$$

where $\Gamma = 1/L$ is the reciprocal inductance, $G = 1/R$ is the conductance, and i_0 is the current at time $t = 0$. The initial conditions i_0 and e_0 in our work will be assumed to be zero.

Mutual inductance arises when two coils are coupled so that some of the magnetic flux is common to both coils. Consider Fig. 1.2 which shows two coils coupled with a mutual inductance M. The dot above a winding is at the side of the coil that becomes positive when the current in the other coil increases in the *assumed* direction. Inductances L_1 and L_2 are the self-inductances of the two coils; L_1 is the inductance of the left coil of Fig. 1.2 when no current flows in the right coil, and L_2 is the inductance of the right coil of Fig. 1.2 when no current flows in the left coil. The basic law of mutual inductance is characterized by the simple loop equations for the two coils

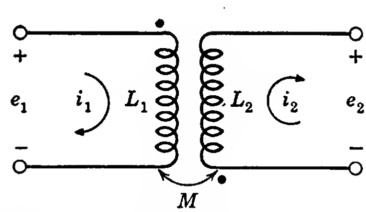


Fig. 1.2. Voltages and currents in a transformer.

$$\begin{aligned} L_1\frac{di_1}{dt} + M\frac{di_2}{dt} &= e_1 \\ M\frac{di_1}{dt} + L_2\frac{di_2}{dt} &= -e_2 \end{aligned} \tag{1.3}$$

It should be clear from these equations that mutual inductance is only a special kind of inductance. Had one of the dots in Fig. 1.2 been reversed, or had one of the two currents been assumed to flow in the direction opposite to that shown (either of which is equivalent to reversing the connections to one coil), the signs of the terms involving M in eqs. 1.3 would be reversed. A handy rule for assigning the sign to a term involving mutual inductance is: use a plus if both the relevant mutual currents flow towards the dots (or both away) and a minus otherwise.

A word as to how the dot locations are determined is probably a worthwhile addition here. Consider again Fig. 1.2. Suppose a battery is placed across L_1 with the positive end at the dot, which can define the dot location as being at either end of the coil. Let this battery suddenly be connected to the coil L_1, as with a switch. Then, an impulse of voltage will appear across L_2. At the end of the coil L_2 that becomes positive, the second dot is placed.

Now let us work backwards from given dot locations, again referring to Fig. 1.2. Suppose we are interested in the sign of the mutual term in the equation for the current i_1. The proper sign is determined with a two-step procedure described as follows. Let a battery be placed across L_2 with the positive side at the dot. The resulting voltage impulse is positive at the dotted end of L_1. If this gives a voltage drop in the i_1 circuit in the assumed direction of i_1, we tentatively put down a positive sign for the mutual term; if it gives a voltage rise in the assumed direction of i_1, we tentatively put down a negative sign. Next we examine the circuit containing i_2. If the battery with the positive end at the dot of coil L_2 causes an actual current that is in the assumed direction of i_2,

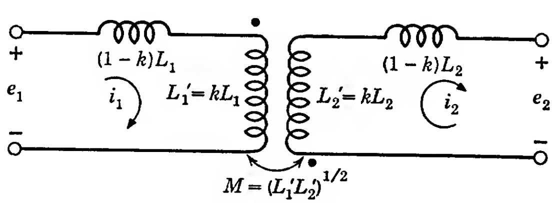


Fig. 1.3. **Leakage inductance transformer equivalent.**

no modification of the tentative sign of the mutual term in the i_1 equation is required. However, if the actual current in L_2 with the positive side of the battery at the dot is opposite to the assumed direction of i_2, we must reverse the tentative sign of the mutual term in the equation for i_1. Since a transformer is a bilateral circuit element, the sign of the mutual term must always be the same in both current equations.

The coefficient of coupling of a transformer is defined as

$$k = M/(L_1L_2)^{1/2} \tag{1.4}$$

which can never be larger than unity.

The transformer of Fig. 1.2 has the equivalent circuit of Fig. 1.3, where $(1 - k)L_1$ and $(1 - k)L_2$ are called the leakage inductances, and where the transformer made up of the inductances L_1' and L_2' has a coefficient of coupling that is unity. The equations for this equivalent circuit are the same as those for Fig. 1.2, which proves it to be an exact equivalent. The reader can easily prove this for himself.

If the two sides of a transformer are connected together in the manner shown by Fig. 1.4*a*, other equivalent circuits consisting of three inductances can be derived. The "T" equivalent is shown in Fig. 1.4*b*, and

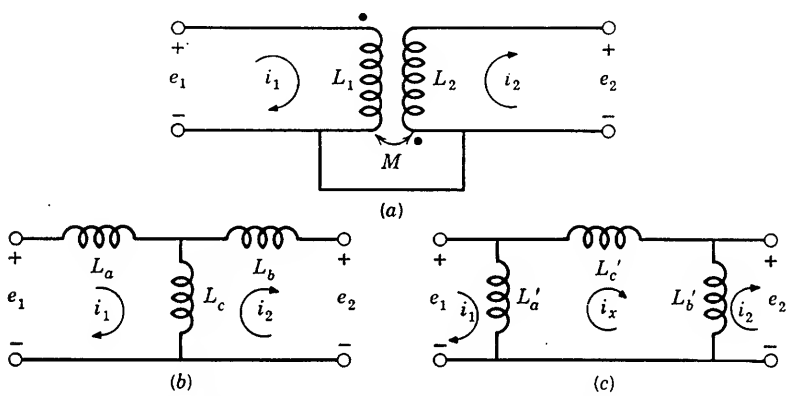


Fig. 1.4. T and pi transformer equivalents.

the "pi" equivalent in Fig. 1.4*c*. For the T equivalent with dots and assumed loop current directions as shown in Fig. 1.4*a*

$$\begin{aligned} L_c &= -M \\ L_a &= L_1 - L_c = L_1 + M \\ L_b &= L_2 - L_c = L_2 + M \end{aligned} \tag{1.5}$$

from which it is to be noted that an equivalent negative inductance appears. For the pi equivalent

$$\begin{aligned} L_1 &= \frac{L_a'(L_b' + L_c')}{L_a' + L_b' + L_c'} & L_a' &= \frac{L_1L_2 - M^2}{L_2 + M} \\ L_2 &= \frac{L_b'(L_a' + L_c')}{L_a' + L_b' + L_c'} & L_b' &= \frac{L_1L_2 - M^2}{L_1 + M} \\ M &= \frac{-L_a'L_b'}{L_a' + L_b' + L_c'} & L_c' &= \frac{L_1L_2 - M^2}{-M} \end{aligned} \tag{1.6}$$

With the dots as shown, L_c' is negative for the pi equivalent and both L_a' and L_b' are positive. To take care of the case when one of the windings of the transformer is reversed, change the sign of M.

The basic unit of inductance L and mutual inductance M is the henry (h). A current change of 1 ampere per second through a 1-h inductor results in a voltage of 1 volt across the inductor. The millihenry (mh) is 10^{-3} h, and the microhenry (μh) is 10^{-6} h. Inverse inductance is not usually specified in a circuit diagram, the value of inductance being given instead.

The basic unit of capacitance is the farad (f). A current of 1 ampere forced to flow into a 1-f capacitor causes the voltage across the capacitor to change at a rate of 1 volt per second. The farad is much too large a unit to be convenient. More practical measures are the microfarad (μf), which is 10^{-6} f, and the micromicrofarad ($\mu\mu$f), which is 10^{-12} f. (The unit $\mu\mu$f is sometimes written "pf" meaning "pico" farads.) Inverse capacitance S is measured in darafs (farad spelled backwards), which is designated df. Megadarafs and megamegadarafs are the more practical measures. Element values are always expressed in farads rather than in darafs.

The basic unit of resistance R is the ohm. On a circuit diagram, a resistance of 10 ohms is commonly written 10 Ω and sometimes simply 10 with the Ω understood. A resistance of 50,000 Ω is commonly written as 50K, where K signifies thousands of ohms. A megohm is 1,000,000 Ω; 10 megohms is commonly designated as 10M. The conductance G of a 1 Ω resistor is 1 mho (℧), which is ohm spelled backwards. One micromho is 10^{-6} ℧, which is a resistance of 1 megohm. Ohms rather than mhos are usually employed in specifying element values.

1.2 Ideal and practical transformers

An often useful artifice is the "ideal" transformer. This device has zero leakage inductance and infinite primary and secondary inductances,

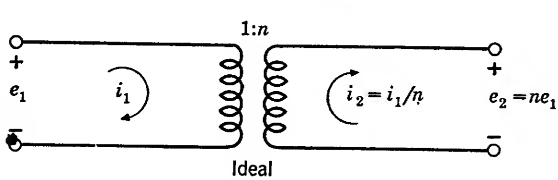


Fig. 1.5. The ideal transformer.

but so proportioned that their ratio is a finite number. The voltage-current relationships in the ideal transformer are shown in Fig. 1.5, where n is the "turns ratio" (a purely real number). The volt-amperes

in primary and secondary are equal, as is proper. However, the ratios of voltage to current differ. Since $e_2/i_2 = n^2e_1/i_1$, an inductance in the secondary has the same effect, as far as the primary is concerned, as an inductance $1/n^2$ times as large placed in the primary. This is demonstrated by the equivalent circuits shown in Fig. 1.6. Resistance R and elastance $S = 1/C$ transform in the same manner as inductance L.

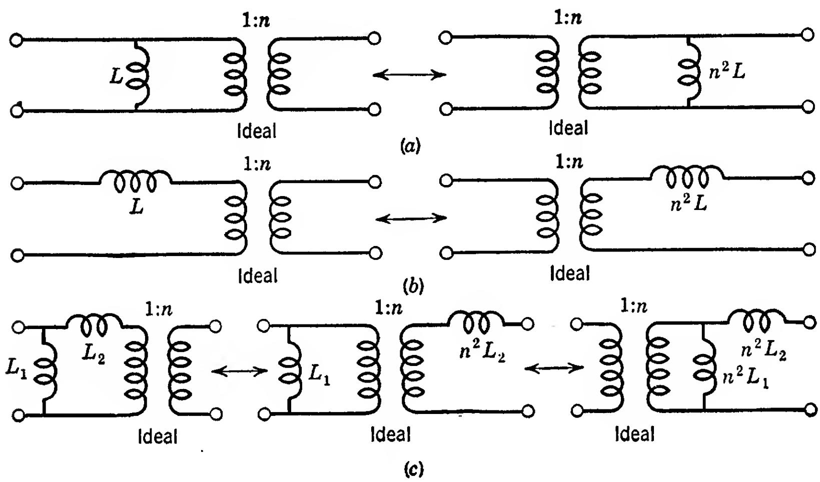


Fig. 1.6. Referring inductances about an ideal transformer.

It must be pointed out that element values transform in this manner only with *ideal* transformers.

Often it is necessary to change the impedance level in some network. This can always be done on paper with an ideal transformer, although it is hardly possible to build such a transformer. (Power transformers are the closest to ideal transformers that exist. Transformers having nonmagnetic cores are not even approximately ideal.) However, if an ideal transformer in a network is adjacent to series and shunt inductances, the ideal transformer and these inductances can be combined into a practical transformer (or its T or pi equivalent). In order to do this, there must be some series inductance to account for the finite leakage inductance of a practical transformer. Also, there must exist some shunt inductance to account for the finite primary and secondary inductances of a practical transformer. The circuit of Fig. 1.6*c* corresponds to a practical transformer.

As an example, consider the circuit of Fig. 1.7. Let us extract the ideal along with nearby inductances and study it separately. In addition, let us "refer" L_2' to the primary side as an inductance $L_2'/n^2 = L_2$.

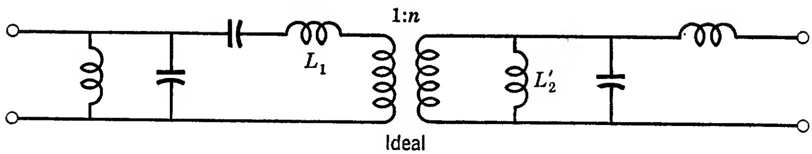


Fig. 1.7. Circuit with ideal transformer.

Then we have the circuit of Fig. 1.8*a* in which terminal conditions have been imposed. The terminal conditions can also be referred to give the circuit of Fig. 1.8*b* (where the dot locations are assumed to be arbitrary),

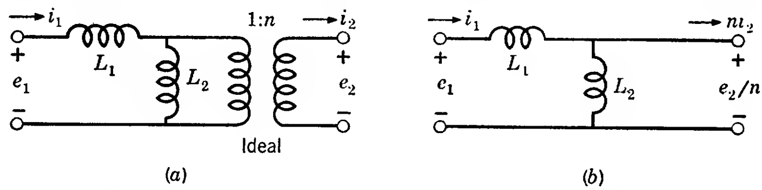


Fig. 1.8. An ideal transformer with associated inductances.

the equations for which are

$$
\begin{aligned}
(L_1 + L_2)\frac{di_1}{dt} - L_2\frac{dni_2}{dt} &= e_1 \\
-L_2\frac{di_1}{dt} + L_2\frac{dni_2}{dt} &= -\frac{e_2}{n}
\end{aligned}
\tag{1.7}
$$

Multiplying the second equation through by n, we get

$$
\begin{aligned}
(L_1 + L_2)\frac{di_1}{dt} - nL_2\frac{di_2}{dt} &= e_1 \\
-nL_2\frac{di_1}{dt} + n^2L_2\frac{di_2}{dt} &= -e_2
\end{aligned}
\tag{1.8}
$$

which are the same as the equations for a transformer with mutual inductance nL_2 and primary and secondary inductances $L_1 + L_2$ and n^2L_2 respectively. Thus, the circuit of Fig. 1.7 can be built as shown in Fig. 1.9. The coefficient of coupling of the resulting practical transformer can easily be found to be $1/(1 + L_1/L_2)^{1/2}$ which, for finite L_1 and L_2, is always less than unity.

Sometimes it is desirable to use a physical T or pi network instead of a transformer. This will not be possible unless all the inductances in the equivalent are positive. If the winding directions of the coils on the

transformer are not important, the mutual element in the equivalent can always be made positive. However, if the primary and secondary inductances of the transformer are considerably different and the coefficient of coupling is not small, one of the inductances of the equivalent T or pi may be negative even though the mutual inductance is positive. If some additional circuit inductance is located adjacent to the objec-

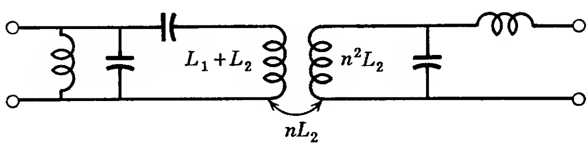


Fig. 1.9. Resulting circuit containing a practical transformer.

tionable negative inductance, it may be enough to cancel the negative inductance and make the equivalent practical.

It has been shown that an ideal transformer in association with finite series and shunt inductances can be converted to a practical transformer which, under certain circumstances, can further be converted to a practical equivalent T or pi network of inductances. In certain cases, there also exists an equivalence between an ideal transformer in association with series and shunt capacitors (or resistors) and a T or pi network of capacitors (or resistors). The method of finding the equivalent is similar to that already described. The details have been left to the problems.

1.3 Loop and node equations

The equations of an electric network may be written in a systematic way using either of Kirchhoff's two laws. The current law states that the algebraic sum of the currents leaving a node (the junction of two or more circuit elements) is zero. The voltage law states that the algebraic sum of the voltage drops around any closed path in a network is zero. In obtaining the equations for a given circuit, one equation must be written for each loop or node depending upon which of Kirchhoff's two laws is employed. The result is a set of simultaneous linear differential equations.

Let a straight line segment with terminals (o——o) represent a *single* circuit element. A generalized network can then be represented by a geometric pattern as, for example, that of Fig. 1.10. If two parts of a network have no branch or node in common, as if the two parts are separated by pure mutual inductance or are completely independent, they are said to be separable. The circuit of Fig. 1.10 has two separate parts. One of the nodes of Fig. 1.10 is termed the reference node and is

indicated as a ground connection. A node in combination with the reference node is called a node pair. A branch is an element or series combination of elements between nodes that connect more than two elements.

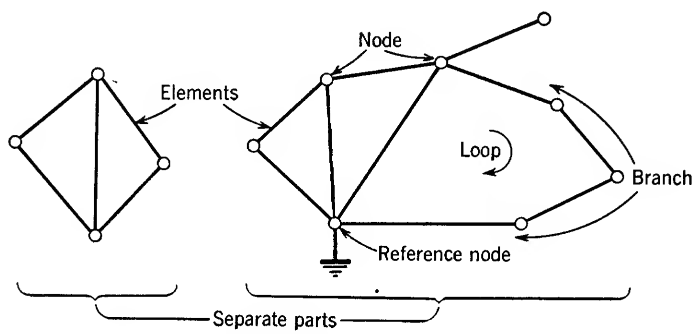


Fig. 1.10. The graph of a network.

The theory of networks when cast in the form of geometric patterns belongs to a branch of mathematics called graph theory. Two important results from this theory are

$$\begin{aligned} \text{Independent loops} &= \text{Elements} - \text{Nodes} + \text{Separate parts} \\ \text{Independent node pairs} &= \text{Nodes} - \text{Separate parts} \end{aligned} \tag{1.9}$$

When a network is to be solved using one of Kirchhoff's laws, it is easiest to use the law which yields the fewest equations. From graph theory, we may determine how many equations will be needed when summing either currents into nodes or voltage drops around closed paths. The required number of loop equations when summing voltage drops is

$$\text{Loop equations} = \text{Independent loops} - \text{Current sources} \tag{1.10}$$

Similarly, the number of nodes at which node currents must be summed is

$$\text{Node equations} = \text{Independent node pairs} - \text{Voltage sources} \tag{1.11}$$

The method of node voltages is often easier in electron-tube circuits because several circuit elements so often appear in shunt (parallel). Also, voltages rather than currents are normally considered the unknowns, which gives a further advantage to the nodal method. When a simple mutual inductance exists, it may often be replaced with its pi or T equivalent to facilitate a node-voltage solution. When several mutual inductances exist, it is more convenient to use loop currents.

Let us use Fig. 1.10 as a specific example for calculating the number of equations required for solving the circuit. There are a total of 15 elements, 12 nodes, and 2 separate parts. Therefore, there are $15 - 12 + 2 = 5$ independent loops and $12 - 2 = 10$ independent node pairs.

It is probably obvious to the reader that voltage generators placed in series with the branches of Fig. 1.10 do not have an effect on the number of loop equations required. Similarly, any number of currents may be injected into the various nodes without affecting the number of node equations required.

In order that the relations of eqs. 1.10 and 1.11 be correct when voltage and current generators are present, it is necessary also to count the generators as elements. In addition, the junction between a voltage generator or a current generator and a single circuit element must be counted as a node.

Two extra rules, the validity of which should be self-evident, can occasionally greatly reduce the apparent number of equations required in a loop analysis. If a network having but a single pair of terminals occurs in shunt with an *ideal* voltage generator, and if the loop currents within this network are not of interest, then the entire network can be ignored. That this can be done follows from the simple fact that the voltage of an ideal source is quite independent of whatever network is connected in parallel with the source. Similarly, if an *ideal* current source occurs in series with some network having but a single pair of terminals, and if the currents within this network are of no interest, then the entire network can be ignored.

It will be noted in eq. 1.9 that the number of independent node pairs is reduced by the number of separate parts. Essentially, this means that a reference node must be chosen for each separate part; otherwise, the separate parts except the one containing the principal reference node would "float." Consequently, when setting up a system for a nodal analysis, the network is made to have but a single separate part by suitably connecting nodes together, for example, by assuming a common reference node as one node in each separate part.

Consider a typical node in some general network. Let the node voltage at this node be e_1, and let the voltages of nearby nodes be e_2, e_3, and e_4, all with respect to the reference node. Assume that driving currents i_1, i_2, i_3, and i_4 are injected into the corresponding nodes, as in Fig. 1.11. A current generator causes the current to flow out of (or into) a node regardless of the node voltage. As a matter of convention, all node voltages, if unknown, are assumed to be positive and are measured with respect to the voltage of the reference node. This has the advantage of simplifying the formulation of the equations.

Kirchhoff's current law states that the algebraic sum of currents leaving (or entering) a node is zero. Let us write the equation for the node having the voltage e_1 in Fig. 1.11. The current leaving through Γ is $\Gamma \int (e_1 - e_2)\, dt$. (If we were summing currents entering the node rather than leaving it, we would write $e_2 - e_1$ rather than $e_1 - e_2$ under the integral.) Similarly, the current leaving through C is $Cd(e_1 - e_3)/dt$.

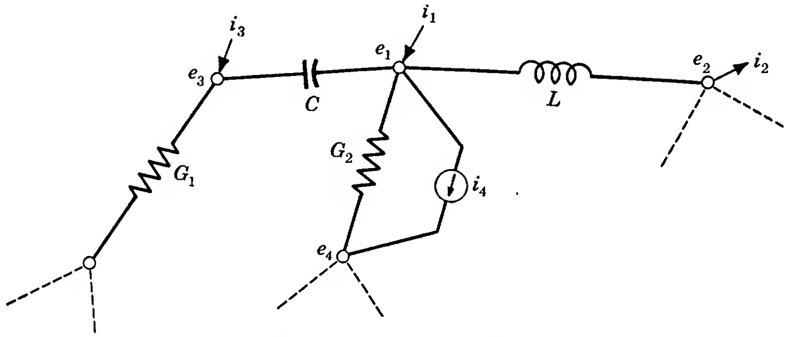


Fig. 1.11. A typical node.

The current i_1 is shown entering the node; hence $-i_1$ leaves. Thus, the equation for the node takes the form

$$\Gamma \int (e_1 - e_2)\, dt + C \frac{d(e_1 - e_3)}{dt} + G_2(e_1 - e_4) + i_4 - i_1 = 0 \quad (1.12)$$

If we collect all terms containing e_1, all containing e_2, and so forth, and put currents i_1 and i_4 on the right side of the equation, we get

$$\left[\Gamma \int e_1\, dt + C \frac{de_1}{dt} + G_2 e_1\right] - [\Gamma \int e_2\, dt] - \left[C \frac{de_3}{dt}\right] - [G_2 e_4] = i_1 - i_4 \quad (1.13)$$

which states in words that the sum of the currents leaving through network elements is equal to the sum of the currents entering from current generators.

From eq. 1.13 a pattern can be seen. The terms multiplying e_1 contain only the elements connected to node e_1; the terms multiplying e_2 contain only the elements common to both e_1 and e_2, and so on. Thus, we could have written eq. 1.13 directly without first going through the step indicated by eq. 1.12. First, the current away from e_1 is calculated assuming all other node voltages to be zero (that is, at the potential of the reference node), which gives the first term of eq. 1.13 and the cur-

rents i_1 and i_4. Next, the current away from node e_1 is calculated assuming all node voltages to be zero except e_2, which gives the second term of eq. 1.13. The third term is obtained by assuming all voltages except e_3 to be zero, and so on.

As a specific example, consider the circuit of Fig. 1.12. There are a total of 12 elements (counting the two generators), 4 nodes, 1 separate part, 2 current generators, and no voltage generators. The number of independent loops is $12 - 4 + 1 = 9$, and the required number of loop equations is $9 - 2 = 7$. The number of independent node pairs is

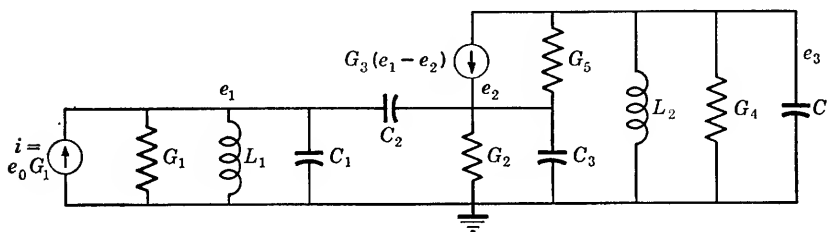


Fig. 1.12. Circuit for the node example.

$4 - 1 = 3$, and the required number of node equations is $3 - 0 = 3$. Evidently, the set of node equations is much simpler in this case. If e_1 is a known voltage (which would be assumed if the ratio e_3/e_1 were of interest, for example), then an ideal voltage generator is placed between the e_1 node and the reference node, and the elements G_1, L_1, and C_1 and the current generator e_0G_1 are all ignored. No equation would then be written for the e_1 node.

The node equations for the circuit of Fig. 1.12 are given in eqs. 1.14. The reader is advised to follow them through carefully and write them down a few times without reference to the text. This example will be used repeatedly throughout the balance of this chapter. The circuit is that of a practical vacuum-tube amplifier.

$$\left[G_1e_1 + \Gamma_1\int e_1\,dt + (C_1 + C_2)\frac{de_1}{dt}\right] - \left[C_2\frac{de_2}{dt}\right] - [0] = i = G_1e_0$$

$$-\left[C_2\frac{de_1}{dt}\right] + \left[(G_2 + G_5)e_2 + (C_2 + C_3)\frac{de_2}{dt}\right] - [G_5e_3] = G_3(e_1 - e_2) \qquad (1.14)$$

$$-[0] - [G_5e_2] + \left[(G_4 + G_5)e_3 + \Gamma_2\int e_3\,dt + C_4\frac{de_3}{dt}\right] = -G_3(e_1 - e_2)$$

To write down a set of node equations without confusion, it is necessary to assume that the polarity of all unknown node voltages is posi-

tive. To write down a set of loop equations, the direction of each unknown loop current must be assumed. By convention, all loop currents are assumed to travel in a clockwise direction. Consider a typical loop in a network as depicted in Fig. 1.13. Voltages e_1 and e_2 result from voltage generators, and i_1, i_2, and i_3 are the assumed loop currents. Let us write the loop equation for current i_1 by following the direction of i_1 around the loop and equating the sum of all the voltage drops to zero.

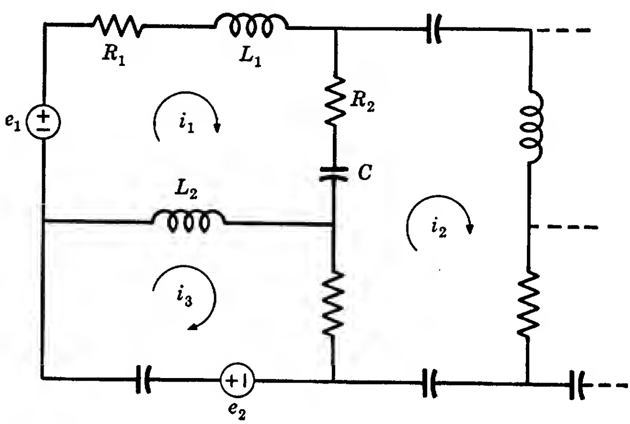


Fig. 1.13. A typical loop.

Since only i_1 flows in R_1 and L_1, the voltage drop in these two elements is $R_1i_1 + L_1\, di_1/dt$. The current $i_1 - i_2$ flows in R_2 and S in the direction of i_1; hence the voltage drop across these two elements is $R_2(i_1 - i_2) + S\int (i_1 - i_2)\, dt$. Similarly, the voltage drop across L_2 is $L_2 d(i_1 - i_3)/dt$. Going through the generator, voltage e_1 is a voltage rise in the direction of i_1; hence the voltage drop is $-e_1$. The equation for loop current i_1 takes the form

$$R_1i_1 + L_1\frac{di_1}{dt} + R_2(i_1 - i_2) + S\int (i_1 - i_2)\, dt + L_2\frac{d(i_1 - i_3)}{dt} - e_1 = 0 \tag{1.15}$$

Collecting terms as we did before

$$\left[(R_1 + R_2)i_1 + (L_1 + L_2)\frac{di_1}{dt} + S\int i_1\, dt\right] - [R_2i_2 + S\int i_2\, dt] - \left[L_2\frac{di_3}{dt}\right] = e_1 \tag{1.16}$$

which in words says that the sum of the voltage drops in circuit elements is equal to the sum of the voltage rises through voltage generators.

This equation could also have been written by inspection. The first term is the sum of the voltage drops in the direction of i_1 obtained when all the currents except i_1 are assumed to be zero; the second term is the voltage drop in the circuit containing i_1 when all the currents except i_2 are assumed to be zero, and so on.

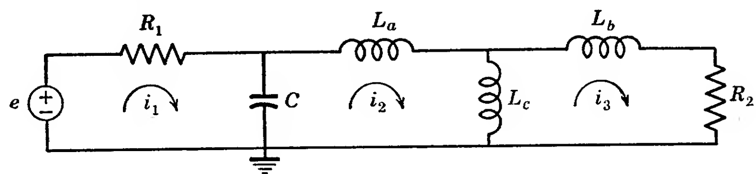


Fig. 1.14. Circuit for the loop example.

As an example, consider the transformer-coupled circuit of Fig. 1.14 in which the equivalent T has been used in place of the transformer. The three loop equations are

$$[R_1 i_1 + S\int i_1\,dt] - [S\int i_2\,dt] - [0] = e$$

$$-[S\int i_1\,dt] + \left[S\int i_2\,dt + (L_a + L_c)\frac{di_2}{dt}\right] - \left[L_c\frac{di_3}{dt}\right] = 0 \quad (1.17)$$

$$-[0] - \left[L_c\frac{di_2}{dt}\right] + \left[R_2 i_3 + (L_b + L_c)\frac{di_3}{dt}\right] = 0$$

As in the previous example, this example represents a practical vacuum-tube circuit and will also be referred to throughout the balance of this chapter.

When a network is to be solved using either loop currents or node voltages, the proper number of currents or voltages as given from graph theory should always be employed. If too many are assumed, the set of equations describing the network will be unnecessarily complex. If too few are assumed, there will not exist enough information to obtain some desired solution; that is, certain variables will be unaccounted for.

In the nodal method, no choice as to the unknown node voltages is given and the analysis procedure is pretty well fixed except for the choice of the reference node. In the loop-current method, however, a choice can often be made as to the exact paths followed by the assumed loop currents. (Nodal analysis is actually simpler than loop analysis for this

reason.) Consider, for example, the circuits of Fig. 1.15. There are eight elements (including the generator), one separate part, and five nodes (including the junction between the generator and resistor); hence, $8 + 1 - 5 = 4$ loop currents are required. The currents in either Fig. 1.15a or b, as well as certain other possible configurations of currents, are satisfactory. However, the currents assumed in Fig. 1.15c are not

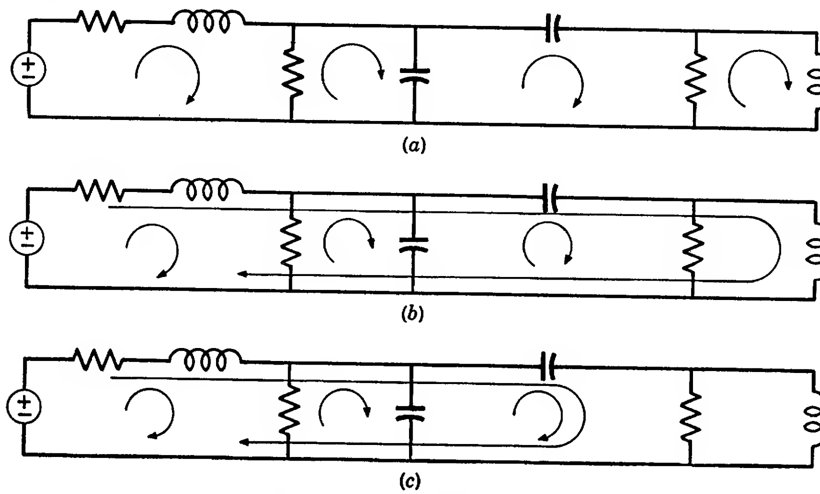


Fig. 1.15. Choice of loop currents.

correct because no current has been assumed to flow in one of the branches. The arrangement of currents generally employed is that in Fig. 1.15a.

The system equations can be made to have slightly different forms depending upon the chosen configuration of loop currents. Often one particular choice leads to simplifications. For example, it is best to have but a single assumed current flowing in the element that represents the output.

We have designated a resistance by R or G in the various figures so far. It really does not matter which symbol is employed; R or G simply designates some particular element. For example, we might designate a 10 Ω resistor as R_1. If we put down G_1 instead of R_1, no difference is intended; a 10 Ω resistance and a 0.1 ℧ conductance are one and the same thing.

1.4 Exchange of sources

When employing the loop-current method of analysis, it is most convenient (although not necessary) to have all sources expressed as voltage

sources, that is, a voltage generator in series with some source network element or elements. For nodal analysis, it is more convenient to have all sources expressed as current sources, that is, a current generator in shunt with some source network element or elements. The exchange is depicted in Fig. 1.16 for source elements of pure R, L, and C with initial currents and voltages i_0 and e_0. The equivalence of the various sources

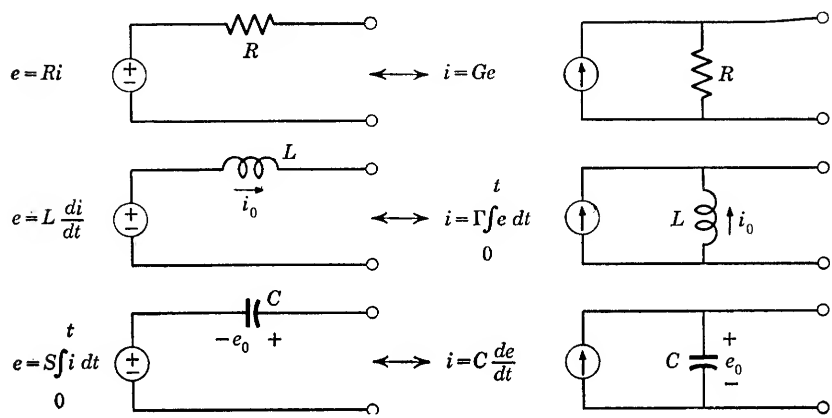


Fig. 1.16. Exchange of sources with initial conditions.

of Fig. 1.16 can readily be demonstrated by writing the pertinent differential equations.

When initial conditions are zero, it is a simple matter to obtain a transformation for a rather general collection of source elements which include those of Fig. 1.16 as special cases. Let some network of elements having two terminals be characterized by an impedance Z or an admittance $Y = 1/Z$. The current source has the network in shunt; the voltage source has the same network in series. The general exchange is depicted in Fig. 1.17. Figure 1.17*c* is essentially the same as Fig. 1.17*a*; it has been included for the purpose of clarifying the manipulation of circuits.

Caution regarding the exchange of sources is required when transient solutions are of interest. Then, initial conditions cannot always be ignored and the general exchange of Fig. 1.17 may not be valid. However, the general exchange is always valid if the initial conditions are zero.

When a voltage generator has zero source impedance, it can be converted to a current generator by assuming a source impedance that can be made arbitrarily small. Should a current generator be ideal in the sense that its shunt impedance is infinite, an arbitrarily large shunt

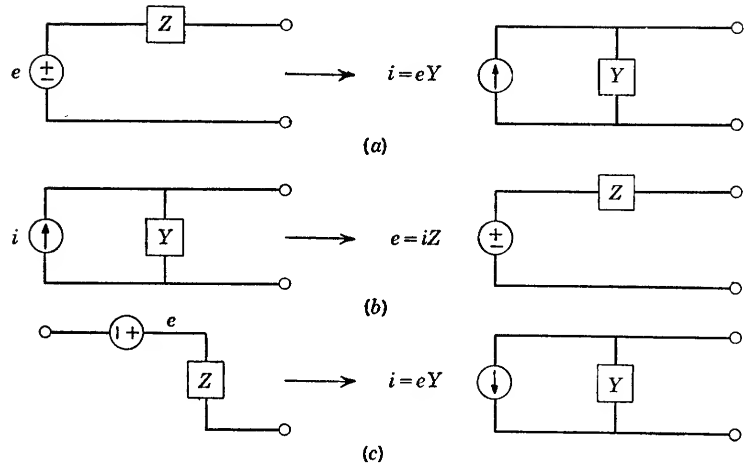


Fig. 1.17. Exchange of sources with zero initial conditions.

impedance can be assumed in order to effect a conversion to a voltage source.

1.5 The steady state and the differential operator

By the steady-state behavior is meant the behavior of circuits containing pure sinusoidal variations of voltage and current that change neither in frequency nor amplitude and which have been going on for all time and will presumably continue to go on for all time. Thus, a voltage or current can be described as $E_m \cos(\omega t + \theta)$ or $I_m \cos(\omega t + \theta)$, where E_m is the peak value of the voltage, I_m is the peak value of the current, θ is a constant phase angle, t is the time, and ω is the frequency in radians per second. The frequency in cycles per second is $f = \omega/2\pi$. *In this book, radian frequency measure will generally be employed.* When talking about devices (as in problems and specifications), frequencies are more often specified in cycles per second (cps), kilocycles per second (kcs), or megacycles per second (mcs).

Circuits act to combine a number of currents and voltages with derivatives and integrals of currents and voltages. Since derivatives and integrals of sine and cosine waves are also sine and cosine waves, the application of a sine wave as a driving function to a network must always result in only sine and cosine variations within the network and at its output, which is but a consequence of the linearity of the network.

Since

$$e^{j\omega t} = \cos \omega t + j \sin \omega t \tag{1.18}$$

where $j = (-1)^{1/2}$ is the "imaginary" number (an "operator"), we can

write

$$\begin{aligned} \cos \omega t &= \operatorname{Re} e^{j\omega t} \\ \sin \omega t &= \operatorname{Im} e^{j\omega t} \end{aligned} \tag{1.19}$$

The symbols Re and Im are called operators because they direct that either a real or an imaginary part, respectively, of a quantity be taken. Consequently, we can use Re $e^{j\omega t}$ instead of $\cos \omega t$ to designate a cosine wave. In fact, we can use simply the exponential itself, $e^{j\omega t}$, tacitly assuming that Re (or Im) is present.

The adoption of the exponential time variable greatly simplifies steady-state relations because

$$\begin{aligned} \frac{d^n e}{dt^n} &= \frac{d^n(E_m e^{j\omega t})}{dt^n} = E_m (j\omega)^n e^{j\omega t} = (j\omega)^n e \\ \int_{n\text{fold}} \cdots \int e \, dt \, dt \cdots dt &= \frac{e}{(j\omega)^n} \end{aligned} \tag{1.20}$$

In other words, taking an nth-order derivative of a voltage or current amounts to multiplication of the voltage or current by $(j\omega)^n$, whereas the process of getting an nth-order integral of a voltage or current amounts to division of the voltage or current by $(j\omega)^n$. All the constants of integration in eq. 1.20 are assumed to be zero in the steady state; hence, the indefinite integrals can be treated without reference to time limits.

The exponential time variable results in a remarkable simplification in that it converts the integrodifferential equations of a system into algebraic equations in $j\omega$ through a *transformation.*

Let us take a somewhat different viewpoint and write derivatives and integrals of voltages and currents in terms of the "derivative operator" p. Then

$$\begin{aligned} \frac{d^n e}{dt^n} &= p^n e \\ \int_{n\text{fold}} \cdots \int e \, dt \, dt \cdots dt &= \frac{e}{p^n} \end{aligned} \tag{1.21}$$

In the steady state, p is simply $j\omega$. However, p has a more general interpretation; in fact, p has no less than four distinct interpretations, most of which we shall treat sooner or later. To avoid a lengthy discussion at this time, we shall consider only $p = j\omega$ in this chapter. The symbol p will be adopted not only because it is more convenient but also because it leads to other interpretations to be explained later.

With the general derivative operator p, a set of network differential equations becomes a set of algebraic equations in p. Setting $p = j\omega$ converts this set of equations to the steady state. In terms of p, eqs. 1.14, which are the equations for the circuit solved on the basis of node voltages given as an example in Sec. 1.3, become

$$\begin{aligned}
&\left[G_1 + \frac{\Gamma_1}{p} + (C_1 + C_2)p\right] e_1 - [C_2 p]e_2 - [0]e_3 = i = G_1 e_0 \\
&-[C_2 p]e_1 + [(G_2 + G_5) + (C_2 + C_3)p]e_2 - [G_5]e_3 = G_3(e_1 - e_2) \\
&-[0]e_1 - [G_5]e_2 + \left[(G_4 + G_5) + \frac{\Gamma_2}{p} + C_4 p\right] e_3 = -G_3(e_1 - e_2)
\end{aligned} \tag{1.22}$$

whereas the equations for the loop-current example, eqs. 1.17, become

$$\begin{aligned}
&\left[R_1 + \frac{S}{p}\right] i_1 - \left[\frac{S}{p}\right] i_2 - [0]i_3 = e \\
&-\left[\frac{S}{p}\right] i_1 + \left[\frac{S}{p} + (L_a + L_c)p\right] i_2 - [L_c p]i_3 = 0 \\
&-[0]i_1 - [L_c p]i_2 + [R_2 + (L_b + L_c)p]i_3 = 0
\end{aligned} \tag{1.23}$$

The techniques for solving sets of algebraic equations with the aid of determinants are well worked out. The reader is advised to review the subject if necessary. For example, currents i_1, i_2, and i_3 for the loop example of eqs. 1.23 can be found in terms of the "driving voltage" e as

$$\begin{aligned}
i_1 &= \frac{+\Delta_{11}{}^m e}{\Delta^m} \\
i_2 &= \frac{-\Delta_{12}{}^m e}{\Delta^m} \\
i_3 &= \frac{+\Delta_{13}{}^m e}{\Delta^m}
\end{aligned} \tag{1.24}$$

where Δ^m signifies the determinant of the coefficients

$$\Delta^m = \begin{vmatrix} \left[R_1 + \frac{S}{p}\right] & \left[-\frac{S}{p}\right] & [0] \\ \left[-\frac{S}{p}\right] & \left[\frac{S}{p} + (L_a + L_c)p\right] & [-L_c p] \\ [0] & [-L_c p] & [R_2 + (L_b + L_c)p] \end{vmatrix} \tag{1.25}$$

and where Δ_{ij}^m is the determinant that remains after striking out the ith row and jth column of the full determinant Δ^m, and is called a minor. The sign to be associated with the determinant Δ_{ij}^m is $(-1)^{i+j}$. We shall consistently use the superscript m (not an exponent) as in Δ_{11}^m to designate a determinant or ratio of determinants or other quantity arising from a loop analysis. The m stands for "mesh," which is a synonym for "loop." The superscript n will be used to denote quantities arising from a nodal analysis.

Equations 1.24 are simply algebraic expressions in the variable p when $p = j\omega$. If the values of the circuit elements are known, the solution for any current involves straightforward manipulations of complex algebra. The usual results of such an analysis are curves of the magnitude and phase angle of a current or voltage as functions of frequency ω.

The numerators and denominators of the currents of eqs. 1.24 will have terms multiplied by p and $1/p$. It is often desirable to remove the factors of $1/p$ by multiplying both numerator and denominator by a suitable power of p so that the currents are expressed as the ratio of two polynomials in p. This can be done in a slightly different and often more convenient manner. Suppose that eqs. 1.17 had been differentiated term by term. This is equivalent to multiplying eqs. 1.23 through by p. The result will be to make the determinant show only powers of p and not $1/p$ so that, when expanded, the desired ratio of polynomials is obtained directly. If we observe that the multiplier of any unknown node voltage or loop current arising from a node or loop analysis, respectively, can never contain other than factors proportional to p, $1/p$, or a constant, the statement that eqs. 1.24 can be put into the form of the ratio of two polynomials in p should be clear. It certainly will become obvious to the reader as he progresses further in this book.

By referring again to the loop example of eqs. 1.23, we can make several useful definitions. The term multiplying i_1 in the first expression is called the self-impedance of the first loop. That multiplying i_2 is the mutual impedance connecting the first and second loops. Let us call z_{ii}^m the self-impedance of the ith loop and z_{ij}^m the impedance common to both the ith and jth loops appearing as a term in the loop equation for the ith assumed current. With this nomenclature, eqs. 1.23 can be written

$$\begin{aligned} z_{11}^m i_1 - z_{12}^m i_2 - z_{13}^m i_3 &= e \\ -z_{21}^m i_1 + z_{22}^m i_2 - z_{23}^m i_3 &= 0 \\ -z_{31}^m i_1 - z_{32}^m i_2 + z_{33}^m i_3 &= 0 \end{aligned} \tag{1.26}$$

which leads to even simpler expressions for the currents given by eqs.

1.24. The determinant is now

$$\Delta^m = \begin{vmatrix} +z_{11}^m & -z_{12}^m & -z_{13}^m \\ -z_{21}^m & +z_{22}^m & -z_{23}^m \\ -z_{31}^m & -z_{32}^m & +z_{33}^m \end{vmatrix} \tag{1.27}$$

A negative sign is associated with all the z_{ij}^m for $i \neq j$. Then for bilateral circuits, all the impedances z_{ij}^m for both $i = j$ and $i \neq j$ will represent physical impedances.

For the example described by eqs. 1.23, z_{13}^m and z_{31}^m are both zero, $z_{12}^m = z_{21}^m$, and $z_{23}^m = z_{32}^m$. For this example, $z_{ij}^m = z_{ji}^m$ for $i \neq j$, which is an indication of a bilateral network in which signals flow with equal ease in either direction. In vacuum-tube circuits, the condition $z_{ij}^m = z_{ji}^m$ is often not true. For example, consider the system of node equations of eqs. 1.22. Combining all factors proportional to the same unknown node voltage, which include two of the currents on the right sides of the equations, eqs. 1.22 become

$$\left[G_1 + \frac{\Gamma_1}{p} + (C_1 + C_2)p\right]e_1 - [C_2p]e_2 - [0]e_3 = i = G_1e_0$$

$$-[C_2p + G_3]e_1 + [(G_2 + G_3 + G_5) + (C_2 + C_3)p]e_2 - [G_5]e_3 = 0 \quad (1.28)$$

$$-[-G_3]e_1 - [(G_3 + G_5)]e_2 + \left[(G_4 + G_5) + \frac{\Gamma_2}{p} + C_4p\right]e_3 = 0$$

In terms of mutual and self-admittances, these equations become

$$\begin{aligned} y_{11}^n e_1 - y_{12}^n e_2 - y_{13}^n e_3 &= i = G_1 e_0 \\ -y_{21}^n e_1 + y_{22}^n e_2 - y_{23}^n e_3 &= 0 \\ -y_{31}^n e_1 - y_{32}^n e_2 + y_{33}^n e_3 &= 0 \end{aligned} \tag{1.29}$$

in which y_{ii}^n is the self-admittance of the ith node and y_{ij}^n is the mutual admittance between the ith and jth nodes appearing as a multiplier of e_j in the equation written for the ith node. Here, it is obvious that $y_{23}^n \neq y_{32}^n$, and so forth. (Although the m and n superscript notation is perhaps a little bulky, it will prevent us from confusing the z's of loop analysis and the y's of node analysis; that is, $z_{11}^m \neq 1/y_{11}^n$, but $z_{11}^m = 1/y_{11}^m$.)

Let us consider the general system of node equations containing a current generator for each node. Such a system can be considered to be

an extension of eqs. 1.29

$$
\begin{aligned}
y_{11}{}^n e_1 - y_{12}{}^n e_2 - \cdots - y_{1r}{}^n e_r &= i_1 \\
-y_{21}{}^n e_1 + y_{22}{}^n e_2 - \cdots - y_{2r}{}^n e_r &= i_2 \\
&\vdots \\
-y_{r1}{}^n e_1 - y_{r2}{}^n e_2 - \cdots + y_{rr}{}^n e_r &= i_r
\end{aligned}
\tag{1.30}
$$

The equivalent general set of equations for loop currents can be obtained from eqs. 1.30 by changing the i's to e's, the e's to i's, and the y's to z's. The i's in eqs. 1.30 remain after *any terms proportional to any of the unknown node voltages have been moved to the left sides of the equations.*

The solution for the node voltage e_s, $s \leq r$, is

$$
e_s = \frac{\sum_{k=1}^{r} \Delta_{ks}{}^n (-1)^{k+s} i_k}{\Delta^n} \tag{1.31}
$$

where Δ^n is the determinant of the admittances. Equation 1.31 is also the solution for the loop current i_s for a system solved on the loop-current basis if the i's and e's are interchanged and the determinants refer to loop impedances.

The summation in the numerator of eq. 1.31 shows the principle of superposition which allows the response of a circuit to several driving functions to be obtained by adding the contributions at the output given by each driving function considered separately.

It should be observed that the ratio of determinants Δ_{ij}/Δ has the dimensions of an impedance or an admittance, depending respectively on whether a node or a loop analysis has been made.

When the differential operator p is equal to $j\omega$ (that is, in the steady state), all impedances and admittances are functions of $j\omega$. We can always split them into real and imaginary parts as

$$
\begin{aligned}
Z(j\omega) &= R(\omega) + jX(\omega) \\
Y(j\omega) &= G(\omega) + jB(\omega)
\end{aligned}
\tag{1.32}
$$

where R and G are called the resistance and conductance respectively, and represent the real part of the impedance or the admittance. X and B are the imaginary parts; $X(\omega)$ is called the reactance (the imag-

inary part of an impedance), and $B(\omega)$ is called the susceptance (the imaginary part of an admittance).

Clearly, the square of the magnitude of $Z(j\omega)$ in eq. 1.32 is

$$|Z(j\omega)|^2 = Z(j\omega)Z^*(j\omega) = Z(j\omega)Z(-j\omega) = [R(\omega)]^2 + [X(\omega)]^2 \quad (1.33)$$

with a similar development for the admittance. Z^* signifies the complex conjugate of Z as

$$Z^*(j\omega) = Z(-j\omega) = R(\omega) - jX(\omega) \quad (1.34)$$

Students are often confused by an expression such as eq. 1.31. Actually, the expression is much simpler than it looks. Suppose, for example, we are interested in solving for e_2 in eqs. 1.30 (which is for $s = 2$ in eq. 1.31). The solution is given simply as the ratio of only two determinants. In the denominator, the full determinant of the admittances appears. In the numerator, the determinant is the same as the full determinant in the denominator except that the second column, which pertains to the voltage e_2, is replaced with the column of currents on the right sides of eqs. 1.30. The summation of eq. 1.31 results simply from the expansion of the determinant in the numerator about the second column, that is, about the column containing the driving currents.

1.6 Input and transfer immittances

The equation giving a node voltage in terms of injected currents has been given in eq. 1.31 and is repeated as

$$e_s = \frac{\sum_{k=1}^{r} \Delta_{ks}{}^n(-1)^{k+s} i_k}{\Delta^n} \quad (1.35)$$

where Δ^n is the determinant of the y's. The corresponding solution for a loop current in terms of driving voltages is

$$i_s = \frac{\sum_{k=1}^{r} \Delta_{ks}{}^m(-1)^{k+s} e_k}{\Delta^m} \quad (1.36)$$

where Δ^m is the determinant of the z's. An "immittance" (a general term signifying either impedance or admittance) relates a voltage and a current according to Ohm's law for a-c circuits. Thus, eqs. 1.35 and 1.36 can be written as the sum of several immittances as

$$\begin{aligned} e_s &= Z_{1s}{}^n i_1 + Z_{2s}{}^n i_2 + \cdots + Z_{rs}{}^n i_r \\ i_s &= Y_{1s}{}^m e_1 + Y_{2s}{}^m e_2 + \cdots + Y_{rs}{}^m e_r \end{aligned} \quad (1.37)$$

The impedance Z_{kk}^n is termed the input impedance of the kth node pair; it is the impedance between the kth node and the reference node relating the voltage at the kth node to the current injected into the kth node from a current generator when all other driving voltages and currents are zero. In reference to a loop analysis, the input admittance Y_{kk}^m is the admittance seen looking into the network between the terminals in the kth loop at the point where the kth driving voltage exists when all other driving voltages and currents are zero. (Causing a driving voltage to become zero essentially short-circuits the voltage source.)

In general, and without reference to either a nodal or a loop analysis, the input immittance of a network relates the voltage and current at one pair of terminals of the network when all other driving currents and voltages are zero.

It must be pointed out that by a driving voltage or current we mean one that is dependent upon an externally applied signal and not one that is dependent upon one of the node voltages or loop currents. For example, the nodal system of Fig. 1.12 has a driving current $i = G_1 e_0$. The other current generator in Fig. 1.12 is entirely dependent upon the node voltages e_1 and e_2 and is always present in the network. It can never be set equal to zero because *it is just as much a part of the network as are the circuit elements themselves.*

In a set of nodal equations, the impedance Z_{jk}^n for $k \neq j$ is a different thing. This impedance relates the voltage appearing at the kth node in response to a driving current at the jth node when all other driving currents and voltages are zero. Since the voltage and current at *different* nodes are related in this manner, Z_{jk}^n is called the "transfer" impedance. With reference to a set of loop equations, the transfer admittance Y_{jk}^m for $j \neq k$ relates the current flowing in the kth loop in response to a voltage applied in the jth loop when all other driving voltages and currents are zero.

Much of the preceding discussion can perhaps be clarified by reference to the examples. Consider first the system of loops of Fig. 1.14. The input admittance of the first loop is

$$i = Y_{11}{}^m e = \frac{+\Delta_{11}{}^m e}{\Delta^m} \tag{1.38}$$

and the transfer admittance relating the third loop current i_3 to e is

$$i_3 = Y_{13}{}^m e = \frac{+\Delta_{13}{}^m e}{\Delta^m} \tag{1.39}$$

where Δ^m is given by eq. 1.25 and where

$$\Delta_{11}{}^m = \begin{vmatrix} +z_{22}{}^m & -z_{23}{}^m \\ -z_{32}{}^m & +z_{33}{}^m \end{vmatrix} \qquad \Delta_{13}{}^m = \begin{vmatrix} -z_{21}{}^m & +z_{22}{}^m \\ -z_{31}{}^m & -z_{32}{}^m \end{vmatrix} \tag{1.40}$$

Now consider the system of node equations defined by Fig. 1.12. The input and transfer impedances relating e_1, e_3, and i are

$$e_1 = Z_{11}{}^n i = \frac{+\Delta_{11}{}^n i}{\Delta^n} \qquad e_3 = Z_{13}{}^n i = \frac{+\Delta_{13}{}^n i}{\Delta^n} \tag{1.41}$$

where Δ^n is the determinant of the y's of eqs. 1.29 and where

$$\Delta_{11}{}^n = \begin{vmatrix} +y_{22}{}^n & -y_{23}{}^n \\ -y_{32}{}^n & +y_{33}{}^n \end{vmatrix} \qquad \Delta_{13}{}^n = \begin{vmatrix} -y_{21}{}^n & +y_{22}{}^n \\ -y_{31}{}^n & -y_{32}{}^n \end{vmatrix} \tag{1.42}$$

Because the full determinant appears in the denominators of all terms of eqs. 1.37, *the denominators of all input and transfer impedances resulting from a nodal analysis are the same.* Similarly, the denominators of all input and transfer admittances resulting from a loop analysis are the same.

The important reciprocity theorem can be deduced from the preceding discussion of input and transfer immittances. If the mutual and self-immittances of a network are symmetric about the main diagonal of the full determinant (that is, if $z_{ij}{}^m = z_{ji}{}^m$ or if $y_{ij}{}^n = y_{ji}{}^n$ for $i \neq j$), the transfer immittances will also be symmetric and cause and effect in the network are interchangeable. This means that a voltage in one loop and a current flowing in another loop are related the same way regardless of which of the two loops contains the voltage and which contains the current. The example we have used for loop currents has this property; the example that has been used for node voltages does not.

Equations 1.39 and 1.41 can be employed to define a more general sort of transfer function which does not (necessarily) have the dimensions of admittance or impedance. Let us first consider the loop system and assume that the output is the voltage that appears across R_2 in Fig. 1.14, e_{out}. Since $e_{\text{out}} = i_3 R_2$, eq. 1.39 becomes

$$e_{\text{out}} = (Y_{13}{}^m R_2)e \tag{1.43}$$

where $Y_{13}{}^m R_2$ is a dimensionless transfer function.

For the node system of Fig. 1.12, the driving current i is indicated as being proportional to a voltage, as $i = G_1 e_0$, which often occurs when a vacuum tube is the current source. Then, assuming e_3 to be the output

with e_0 the input, eq. 1.41 becomes

$$e_3 = (Z_{13}{}^n G_1)e_0 \tag{1.44}$$

which is also a dimensionless transfer function.

When only one driving current or voltage exists, all terms but one on the right of eqs. 1.37 are zero. For convenience, the driving current or voltage can be assumed to occur in the first node or loop and the output at the rth node or loop. Then, the transfer and input immittances are

$$\begin{aligned} e_1/i &= Z_{11}{}^n \qquad e_r/i = Z_{1r}{}^n \quad \text{(node)} \\ i_1/e &= Y_{11}{}^m \qquad i_r/e = Y_{1r}{}^m \quad \text{(loop)} \end{aligned} \tag{1.45}$$

where i and e are the driving current and voltage respectively. If the driving current i is proportional to an applied voltage in a node system, or if the output voltage is proportional to the rth loop current, the transfer function can be expressed as the ratio of two voltages (or two currents) and will be dimensionless.

The input immittance describes the input characteristics of a network. The transfer function relates output and input and is the quantity of central importance (although in vacuum-tube systems a transfer function is often defined in terms of an input immittance). When the steady-state condition exists, the z's and y's of the determinant of a network are functions of $j\omega$ and $j\omega$ only; hence, the transfer and input immittances and the transfer function will be functions of $j\omega$ only. Thus, any network is completely characterized in the steady state by functions of $j\omega$ or the differential operator p when $p = j\omega$.

The transfer function as described up to now relates voltages and/or currents. However, it may relate many different kinds of quantities: voltages, currents, mechanical positions or angles, velocities, fluid flow rates, and so forth.

As defined, the transfer function always means the ratio of the output divided by the input and *never* the input divided by the output.

1.7 Circuit theorems

Several theorems regarding circuits find extensive use in facilitating analysis and synthesis. Some of them have already been mentioned and essentially proved.

The theorem of superposition is a consequence of linearity and really is more a condition than a theorem. Since the output of a network is made up of a linear sum of terms, each of which is dependent upon a single source, the contributions at any point in a network resulting from each source can be added independently. In calculating a given con-

tribution, all driving voltages and currents are assumed to be zero except, of course, the voltage or current whose contribution is sought. Superposition is a condition rather than a theorem because *all* linear systems relating to *all* physical phenomena behave the same way in this respect.

The exchange of sources is properly a theorem. It is an especially useful one in vacuum-tube circuits.

The theorem of reciprocity is applicable to bilateral networks in the steady state. If a voltage source e in branch a of a network produces a

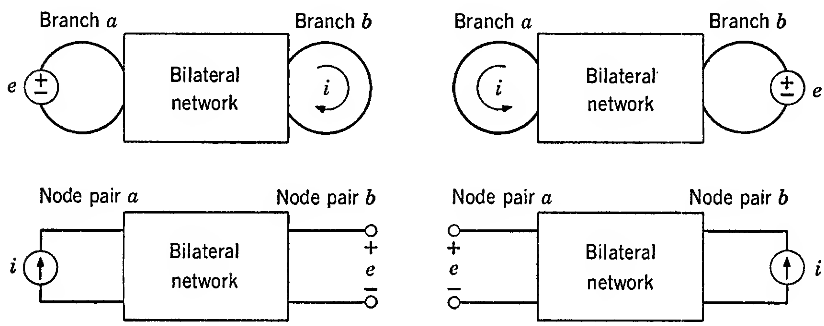


Fig. 1.18. Demonstration of reciprocity.

current i in branch b, the source can be removed from branch a with the gap created by its removal short-circuited and placed in series with branch b. The same current that originally flowed in branch b will now flow in branch a. A similar transfer of current sources between nodes can be effected without changing the node voltages involved. This theorem is demonstrated pictorially in Fig. 1.18.

Two important theorems remaining to be discussed are Thévenin's (and Norton's) theorem and the theorem of maximum power transfer. The proof of these two theorems will be given here because of the mathematical techniques they illustrate.

Consider a "black box" containing any arrangement of voltage and current generators in an arbitrary network. In order to prove Thévenin's theorem, it is only necessary that the complete network obey the law of superposition (that is, that it be linear) and that it be operating in the steady state with all voltage and current generators at the same frequency. Assume that all voltage sources in the black box (see Fig. 1.19) are converted to current sources. The question of the conversion of a voltage source having a zero source impedance can be resolved by assuming a source impedance ϵ where ϵ is arbitrarily small. (Actually,

no practical voltage source can have a source impedance that is precisely zero.)

Assume that the network containing r current sources drives an arbitrary load impedance Z_L placed between the rth node and the reference

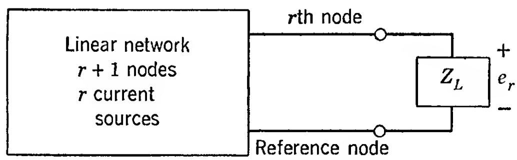


Fig. 1.19. A general network containing current sources.

node. The voltage of the rth node is e_r, which is given by

$$e_r = \pm \frac{\Delta_{1r}{}^n i_1}{\Delta^n} \mp \frac{\Delta_{2r}{}^n i_2}{\Delta^n} \pm \cdots + \frac{\Delta_{rr}{}^n i_r}{\Delta^n} \tag{1.46}$$

where the sign depends upon whether $1 + r$ is odd or even.

The determinant Δ^n is the general determinant of the y's. The self-admittance of the rth node is $y_{rr}{}^n + Y_L$, where $y_{rr}{}^n$ is the admittance of all elements *except* Y_L. Thus

$$\Delta^n = \begin{vmatrix} +y_{11}{}^n & -y_{12}{}^n & -\cdots & -y_{1r}{}^n \\ -y_{21}{}^n & +y_{22}{}^n & -\cdots & -y_{2r}{}^n \\ \vdots & & & \\ -y_{r1}{}^n & -y_{r2}{}^n & -\cdots & +(y_{rr}{}^n + Y_L) \end{vmatrix} \tag{1.47}$$

If $\Delta_0{}^n$ is the value of Δ^n with $Y_L = 0$ (load removed), it can be observed that

$$\Delta^n = \Delta_0{}^n + Y_L \Delta_{rr}{}^n \tag{1.48}$$

where $\Delta_{rr}{}^n$ is the determinant Δ^n with the rth row and the rth column removed. All determinants of the form $\Delta_{jr}{}^n$ are *independent* of Y_L.

Using eq. 1.48 in eq. 1.46, we get

$$e_r = \left(\frac{\pm \Delta_{1r}{}^n i_1 \mp \Delta_{2r}{}^n i_2 \pm \cdots + \Delta_{rr}{}^n i_r}{\Delta_0{}^n} \right) \left(\frac{\Delta_0{}^n}{\Delta_0{}^n + Y_L \Delta_{rr}{}^n} \right) \tag{1.49}$$

The first quantity in parenthesis in eq. 1.49 is the voltage at the terminals of the network with $Y_L = 0$; that is, it is the open-circuit

voltage e'. The second term can be written

$$\frac{\Delta_0{}^n}{\Delta_0{}^n + Y_L\Delta_{rr}{}^n} = \frac{Z_L}{Z_L + \Delta_{rr}{}^n/\Delta_0{}^n} \tag{1.50}$$

The ratio $\Delta_{rr}{}^n/\Delta_0{}^n$ is the input impedance of the rth node with $Y_L = 0$; that is, it is the input impedance to the black box with all driving currents zero and can be called Z_{in}. Thus, it should be clear that the black box has the load voltage

$$e_r = \frac{Z_L}{Z_L + Z_{\text{in}}} e' \tag{1.51}$$

which is equivalent to the circuit of Fig. 1.20.

Thévenin's theorem can well be restated as follows: a linear network having a pair of terminals and containing some combination of voltage

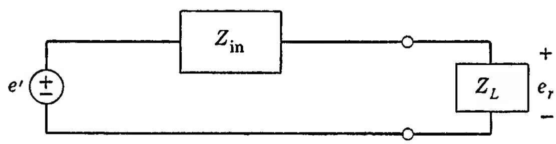


Fig. 1.20. Thévenin equivalent.

and current sources (all sine waves of the same frequency) can be replaced as far as conditions at the terminals are concerned with a single voltage generator in series with a single impedance. The voltage generator has a value equal to the open-circuit voltage at the terminals of the network, and the impedance is that looking back into the network (with the load on the terminals removed) when all driving voltages and currents in the network are zero.

If e' and Z_{in} in Fig. 1.20 are converted to a current source and source admittance, a similar theorem, called Norton's theorem (which is not essentially different from Thévenin's theorem), is obtained.

Figure 1.20 describes a very useful short cut which can be called the "voltage-divider" rule. Equation 1.51 gives the steady-state voltage across an arbitrary impedance when a voltage of a known value is applied to a series combination of impedances.

Thévenin's theorem is often extremely useful in simplifying networks. Frequently, a complicated impedance Z_{in} and voltage e' (where both may be functions of frequency in general) as in Fig. 1.20 can be approximated with much simpler functions over a sufficient band of frequencies to satisfy the requirements of some given problem. The impedance Z_{in} will not necessarily conform to a physically meaningful network.

Another theorem of considerable importance concerns the maximum power transfer in the steady state between a generator and a load. Consider a quite general source and source impedance driving a load as shown in Fig. 1.21. It is assumed that operation is at a fixed frequency. It is desired to find the condition on Z_L for maximum power dissipation

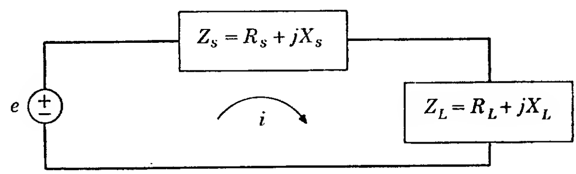


Fig. 1.21. A source and load in the steady state.

in the load. Let i and e be a current and a voltage represented in terms of exponential time variations. Then the power in the load is

$$P_L = |i|^2 R_L/2 = ii^* R_L/2 \tag{1.52}$$

where the use of the complex conjugate should be observed. The current is given by

$$i = \frac{e}{R_s + R_L + j(X_s + X_L)} \tag{1.53}$$

Substituting eq. 1.53 in 1.52

$$P_L = \frac{ee^* R_L/2}{(R_s + R_L)^2 + (X_s + X_L)^2} \tag{1.54}$$

The maximum value of P_L, a function of R_L and X_L, can be established by setting the derivatives of P_L equal to zero

$$dP_L/dX_L = dP_L/dR_L = 0 \tag{1.55}$$

This manipulation yields a maximum power $|e|^2/8R_L$ (where $|e|$ is the peak value of the voltage) when

$$R_L = R_s \quad \text{and} \quad X_L = -X_s \tag{1.56}$$

which means that, for maximum power transfer, the source and load impedances must be complex conjugates. Of course, adjusting the load to absorb maximum power does not yield maximum load voltage.

1.8 Three kinds of networks

A general network can be considered to be a black box having a number of accessible terminals. Driving functions may be applied to some

of these terminals with outputs appearing at other terminals. The box will consist of either a bilateral, unilateral, or unstable network.

A bilateral network must obey the law of reciprocity. Such networks generally consist only of R, C, L, and M with no vacuum tubes. In a bilateral network, signals travel between pairs of terminals with equal ease in either direction. Sometimes, vacuum-tube networks may be operated as two-terminal devices so that they behave like R, L, or C as far as one pair of terminals is concerned (with perhaps a direct voltage or current source as well). When occurring in this manner, a vacuum tube may exist in a network without preventing it from being bilateral (for example, when a vacuum tube with Miller effect is utilized to obtain a controlled input impedance). The determinant of a bilateral network will always display symmetry about the main diagonal; that is, $z_{ij}{}^m = z_{ji}{}^m$ or $y_{ij}{}^n = y_{ji}{}^n$ for $i \neq j$. (Of course, symmetry about the main diagonal may not be evident if the set of equations describing the network is not written in an orderly manner.) Often, a bilateral network having two pairs of terminals is purposely made physically symmetric about a center line so that turning the network end for end introduces no discernible change in the circuit diagram. Then, in addition to displaying symmetry about the main diagonal, the determinant of the network (assuming that loop currents or node voltages are also symmetrically disposed) will also show symmetry about the opposite diagonal.

A unilateral network does not transmit signals with equal ease in either direction between pairs of terminals. The determinant of such a network will never show perfect symmetry about the main diagonal; that is, all the $y_{ij}{}^n$ or $z_{ij}{}^m$ are not equal to the $y_{ji}{}^n$ or $z_{ji}{}^m$. It should be clear that a unilateral network will always contain vacuum tubes (or comparable devices such as transistors) since it is the vacuum tube that is the unilateral element.

An unstable network is one that yields an output without an input being applied. Sine-wave oscillators belong to the class of unstable networks. All unstable networks contain vacuum tubes or comparable devices and are, of course, far from bilateral.

The class of unilateral networks (one of the three main groups of networks) contains two members which have somewhat different basic properties. The simplest unilateral network consists of a cascade of bilateral networks isolated from one another by vacuum tubes which act as perfect coupling devices. Then the over-all transfer function is the product of the transfer functions of the individual bilateral networks. This is described more clearly in Fig. 1.22 for voltage generators, where k_1 and k_2 are constants. The input to the first bilateral network e_1 is modified by the transfer function to give e_2. Then, voltage

e_2 is applied as an input to the second bilateral network after multiplication by a number k_1 and so on. The individual networks have no reaction upon one another. Clearly, the over-all transfer function giving the output to input ratio for the complete system of Fig. 1.22 is

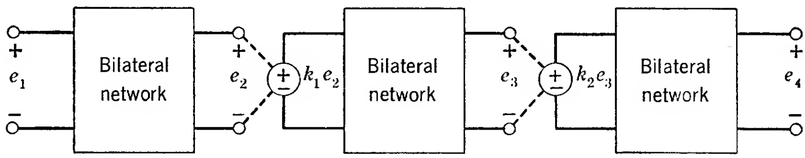


Fig. 1.22. A cascaded system.

simply the product of the transfer functions of the individual "stages"

$$e_4/e_1 = F_1F_2F_3 \tag{1.57}$$

where

$$F_1 = e_2/e_1, \qquad F_2 = e_3/e_2, \qquad F_3 = e_4/e_3 \tag{1.58}$$

Constants k_1 and k_2 relate to the gain given by the vacuum tubes. Consider a typical stage as drawn in Fig. 1.23. The set of loop or node equations for this circuit will have an appearance that is the same as that of a bilateral network, yet we know that the presence of the vacuum tube prevents the system from being bilateral. We shall call such

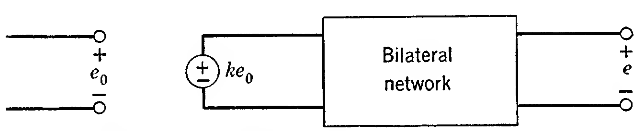


Fig. 1.23. A single stage in a cascaded system.

unilateral networks "isolated" networks. It can be concluded that isolated networks contain vacuum tubes which act as perfect coupling devices between bilateral networks. The example given to describe the loop method of analysis is an isolated network if it is assumed that the voltage generator is a vacuum tube with the input signal applied to the grid of the tube.

The second kind of unilateral network will be called a "feedback" network and will involve tubes which operate in a more complex manner. The over-all transfer function of a feedback network cannot be reduced to the product of transfer functions of individual bilateral networks (unless some different but equivalent system is hypothesized). The determinant of the simplest possible feedback network will not

show symmetry about the main diagonal. *If any of the voltages of voltage generators or currents from current generators are partially or completely dependent upon any of the unknown node voltages or loop currents, the network will be a feedback network.* The example we used to describe nodal analysis is a feedback network. Unstable networks are feedback networks carried to an extreme.

Other types of linear unilateral networks, which utilize unusual materials in the construction of circuit elements, will not be discussed. (An example is the microwave "gyrator.")

In vacuum-tube systems, input, output, and interstage networks consisting of R, L, C, and M are bilateral networks when viewed separately from the vacuum tubes. A conventional amplifying system without feedback is an isolated network. Circuits such as cathode followers, feedback pairs, and servomechanisms are feedback networks. Oscillators, multivibrators, and other waveform generating systems are unstable networks.

Two more definitions associated with feedback networks are worthwhile additions here. In most systems, some input signal traverses a more or less regular path through a vacuum-tube network to arrive finally at the output. "Degenerative" feedback is given by a connection causing a signal at one point in a system to oppose itself at some earlier point. Voltage feedback derives the feedback signal from a voltage, whereas current feedback derives the feedback quantity from a current. "Regenerative" feedback is just the opposite of degenerative feedback; it is given by a connection causing the signal at one point in a system to aid itself at an earlier point. Unstable networks utilize a great amount of regenerative feedback. It is possible for one given system to provide degenerative and regenerative feedback for signals of different frequencies.

1.9 From equations to networks

Consider a generalized system of equations written for a *bilateral* network having a current generator and three independent node pairs.

$$\begin{aligned} +y_{11}{}^n e_1 - y_{12}{}^n e_2 - y_{13}{}^n e_3 &= i \\ -y_{21}{}^n e_1 + y_{22}{}^n e_2 - y_{23}{}^n e_3 &= 0 \\ -y_{31}{}^n e_1 - y_{32}{}^n e_2 + y_{33}{}^n e_3 &= 0 \end{aligned} \tag{1.59}$$

Let us draw a diagram that typifies these equations, using boxes to represent the various self- and mutual admittances. The diagram can be started by assuming three nodes and a reference node as indicated in Fig. 1.24 with the driving current into the first node. The admittance

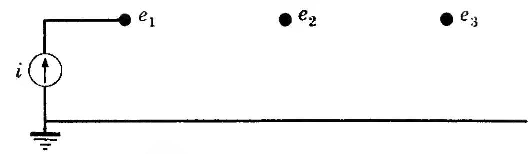


Fig. 1.24. Node placement.

$y_{12}^n = y_{21}^n$ exists between nodes e_1 and e_2, $y_{13}^n = y_{31}^n$ between e_1 and e_3, and $y_{23}^n = y_{32}^n$ between e_2 and e_3. Thus, we can draw a more complete circuit as shown in Fig. 1.25. The self-admittance of a node is

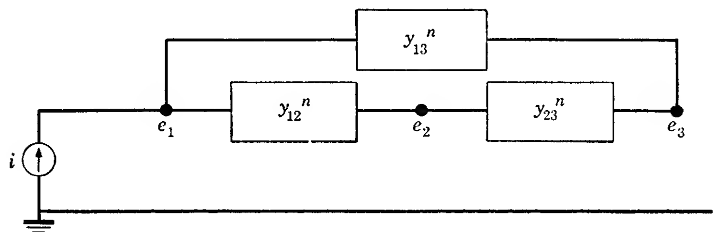


Fig. 1.25. Placement of mutual admittances.

only partly made up of mutual admittances. We can draw the complete network by adding additional self-admittances as shown in Fig. 1.26.

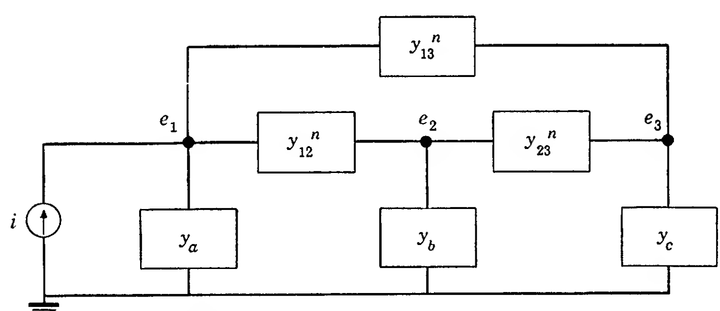


Fig. 1.26. The complete nodal network.

Clearly, it is necessary that

$$
\begin{aligned}
y_a &= y_{11}^n - y_{12}^n - y_{13}^n \\
y_b &= y_{22}^n - y_{23}^n - y_{21}^n \\
y_c &= y_{33}^n - y_{32}^n - y_{31}^n
\end{aligned}
\tag{1.60}
$$

A similar procedure can be followed with a set of loop equations. For example, eqs. 1.61 can be treated by assuming three loop currents with

the driving voltage e in the first loop

$$\begin{aligned} +z_{11}{}^m i_1 - z_{12}{}^m i_2 - z_{13}{}^m i_3 &= e \\ -z_{21}{}^m i_1 + z_{22}{}^m i_2 - z_{23}{}^m i_3 &= 0 \\ -z_{31}{}^m i_1 - z_{32}{}^m i_2 + z_{33}{}^m i_3 &= 0 \end{aligned} \tag{1.61}$$

The result is the network shown in Fig. 1.27, where it is necessary that

$$\begin{aligned} z_a &= z_{11}{}^m - z_{12}{}^m - z_{13}{}^m \\ z_b &= z_{22}{}^m - z_{21}{}^m - z_{23}{}^m \\ z_c &= z_{33}{}^m - z_{32}{}^m - z_{31}{}^m \end{aligned} \tag{1.62}$$

Of course, if the y's or z's given by eqs. 1.60 or 1.62 do not correspond to meaningful immittances, the network cannot be physically constructed.

The reader should observe that system equations written on either the loop or the node basis have the same form; all that need be done to convert a set of equations is to interchange i's and e's and y's and z's. The similarities between the equations are so marked that a special pair of networks called "duals" can be defined. Another important pair of networks is called "reciprocal." If a network solved on the basis of loop currents is known, its dual solved on the basis of node voltages can often be determined, or conversely. However, a dual network might not exist; it will not exist if the network cannot be drawn without crossing lines (in which case, the network is said to be nonplanar).

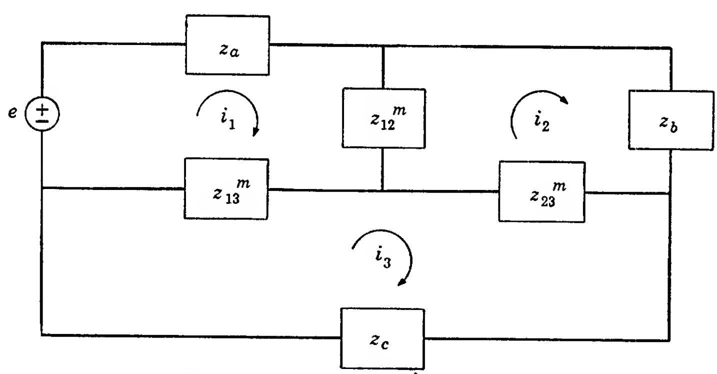


Fig. 1.27. The complete loop network.

1.10 Dual and reciprocal networks

If impedances Z_1 and Z_2 obey the relationship

$$Z_1 Z_2 = R_0^2 \tag{1.63}$$

then the two impedances are said to be reciprocal with respect to the resistance R_0. For example, a series-resonant circuit has the impedance

$$Z_1 = R_1 + pL_1 + 1/pC_1 \tag{1.64}$$

whereas its reciprocal has the admittance

$$Y_2 = Z_1/R_0^2 = G_2 + pC_2 + 1/pL_2 \tag{1.65}$$

which is the admittance of a parallel-resonant circuit with a conductance R_1/R_0^2, a capacitance $C_2 = L_1/R_0^2$, and an inductance $L_2 = C_1 R_0^2$.

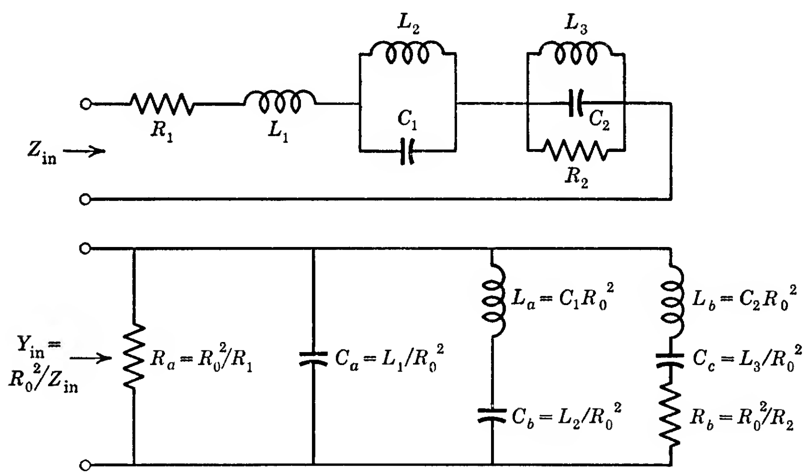


Fig. 1.28. An example of reciprocal impedances.

If a circuit consists of several impedances in series, the total impedance is $Z_1 = Z_a + Z_b + Z_c + \cdots$. The reciprocal has the admittance

$$Y_2 = Z_1/R_0^2 = Z_a/R_0^2 + Z_b/R_0^2 + \cdots \tag{1.66}$$

which is the admittance of a number of admittances in parallel. As an example, consider the reciprocal impedances of Fig. 1.28.

As a further and most practical example, consider the ladder network of Fig. 1.29 driven with a voltage generator of value e_0. The ratio e_0/i_1

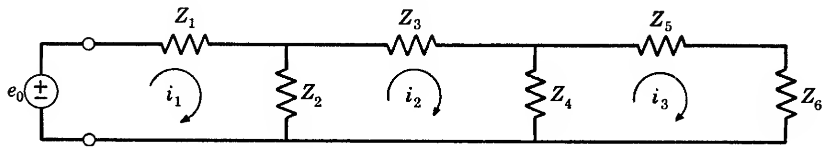


Fig. 1.29. A ladder network.

is the input impedance to the network

$$i_1/e_0 = \Delta_{11}{}^m/\Delta^m = Y_{11}{}^m \tag{1.67}$$

Now assume that a different network is driven with a current i_0 at the first node. The ratio i_0/e_1 is the input admittance of this new network described by

$$e_1/i_0 = \Delta_{11}{}^n/\Delta^n = Z_{11}{}^n \tag{1.68}$$

Let us choose the new network such that the input impedance $Z_{11}{}^n$ will be reciprocal to $Z_{11}{}^m$ with respect to a resistance R_0. Then, $Z_{11}{}^n Z_{11}{}^m = R_0{}^2$, which requires that

$$\Delta_{11}{}^n/\Delta^n = R_0{}^2\Delta_{11}{}^m/\Delta^m \tag{1.69}$$

If all the elements in a row of a determinant are multiplied by a constant, the value of the determinant is multiplied by that constant. Thus, if we form new determinants $(\Delta_{11}{}^m)_N$ and $(\Delta^m)_N$ by multiplying all the rows (or columns) of each by $G_0{}^2$ (or all the rows by G_0 *and* all the columns by G_0), we have

$$\frac{(\Delta_{11}{}^m)_N}{(\Delta^m)_N} = R_0{}^2\,\frac{\Delta_{11}{}^m}{\Delta^m} = \frac{\Delta_{11}{}^n}{\Delta^n} \tag{1.70}$$

where it should be observed that $(\Delta^m)_N$ has one more row and one more column than $(\Delta_{11}{}^m)_N$.

If eq. 1.70 is to be valid, every element in the determinant $(\Delta^m)_N$ must be equal to the corresponding element in Δ^n. Although this is not the only way the equivalence can be obtained (in general, a network function can be realized with several physically different networks), it does represent one important way. Then

$$G_0{}^2 z_{ij}{}^m = y_{ij}{}^n \tag{1.71}$$

From this, we can see that the network of Fig. 1.30 has an input impedance that is reciprocal with respect to that of the network of Fig.

1.29. Each series arm of the original ladder has become a shunt arm, and conversely, and each impedance in the new network is reciprocal to the corresponding impedance in the original ladder.

Consider now a bilateral network having an input pair of terminals and an output pair of terminals as in Fig. 1.31*a*. At the input is placed a voltage generator with a source resistance R_s. The output voltage is

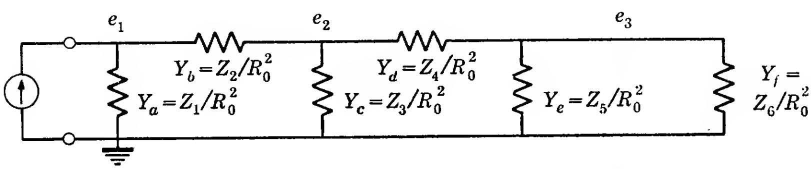


Fig. 1.30. The reciprocal to the ladder network.

$i_k R_L$. Another network driven with a current source and having different source and load resistances (to be determined) is shown in Fig. 1.31*b*. Assume that network 1 has k loops whereas network 2 has k node pairs; that is, the first network has the same number of independent loops as the second has independent node pairs. Then, when the

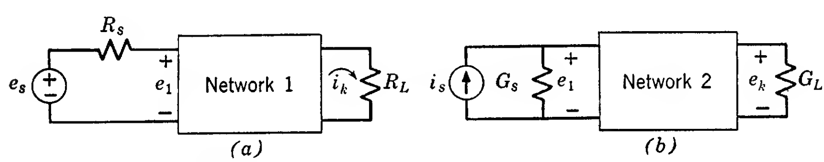


Fig. 1.31. Nomenclature for dual networks.

output is at the kth node or loop with the input at the first node or loop, the transfer functions are

$$\begin{aligned} &\text{Network 1:}\quad i_k/e_s = (-1)^{1+k}\Delta_{1k}{}^m/\Delta^m = Y_{1k}{}^m \\ &\text{Network 2:}\quad e_k/i_s = (-1)^{1+k}\Delta_{1k}{}^n/\Delta^n = Z_{1k}{}^n \end{aligned} \tag{1.72}$$

If $Z_{1k}{}^n Z_{1k}{}^m = R_0{}^2$, the two networks are said to be duals, and the transfer functions of the two networks will be identical except for a constant factor. In order that the networks be duals, we must have

$$\frac{\Delta_{1k}{}^n}{\Delta^n} = R_0{}^2\,\frac{\Delta_{1k}{}^m}{\Delta^m} = \frac{(\Delta_{1k}{}^m)_N}{(\Delta^m)_N} \tag{1.73}$$

where $(\Delta^m)_N$ is the determinant obtained by multiplying the elements in all the rows of Δ^m by $G_0{}^2$ and $(\Delta_{1k}{}^m)_N$ is that formed by multiplying the elements in all the rows of $\Delta_{1k}{}^m$ by $G_0{}^2$.

Equation 1.73 can be satisfied in one way by requiring that each term of $(\Delta^m)_N$ be equal to the corresponding term of Δ^n. This requires

$$z_{ij}{}^m G_0{}^2 = y_{ij}{}^n \tag{1.74}$$

from which it can be seen that the mutual and self-immittances of the two networks are reciprocal. Equation 1.74 is the same relation that resulted in defining ladder networks having reciprocal input impedances.

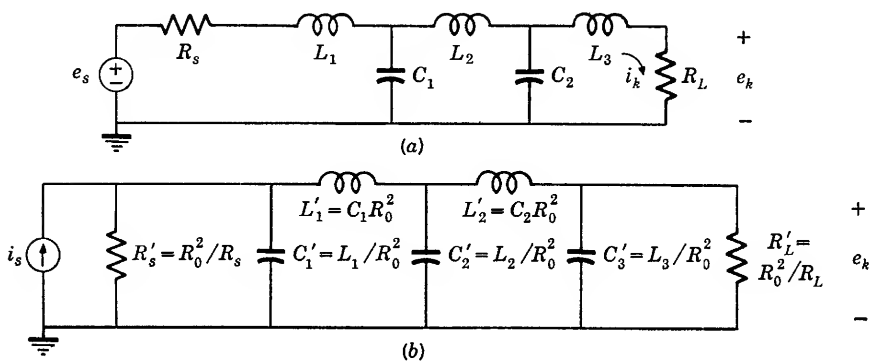


Fig. 1.32. Example of dual networks.

Thus, the dual networks of Fig. 1.32 can be derived as the ladder networks having reciprocal input impedances.

The transfer functions of dual networks, as given by eqs. 1.72 and the relationship $Z_{1k}{}^m Z_{1k}{}^n = R_0{}^2$, are proportional to one another so that the behavior of the two networks with frequency is the same except for a constant of proportionality.

When R_s and R_L are well-defined source and load resistances, it is often convenient to use the particular resistance

$$R_0{}^2 = R_s R_L \tag{1.75}$$

which causes the circuit of Fig. 1.32*b* to become that of Fig. 1.33*a* in which the source and load resistances have been interchanged. By the theorem of reciprocity, the network of Fig. 1.33*a* can be turned end for end without changing the transfer function. If this is done, the circuit of Fig. 1.33*b* is obtained. From eqs. 1.72 (and reciprocity), the transfer function of the network of Fig. 1.33*b* is $e_k/i_s = Z_{1k}{}^n = R_s R_L Y_{1k}{}^m$. If it is assumed that $i_s R_s = e_s$ is the driving voltage, then the current generator of Fig. 1.33*b* can be converted to a voltage generator to give the circuit of Fig. 1.33*c*, for which the voltage transfer function is $e_k/e_s = R_L Y_{1k}{}^m$.

The transfer function of the network with which we started (Fig. 1.32a) is $i_k/e_s = Y_{1k}{}^m$. Since the current i_k flows through the load R_L, the load voltage is $e_k = i_k R_L$. Therefore, the voltage transfer function of the circuit of Fig. 1.32a is *identical* to that of Fig. 1.33c.

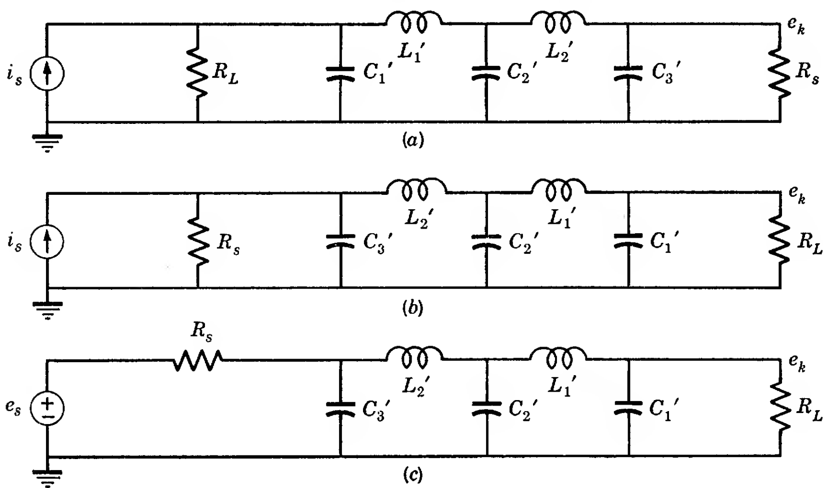


Fig. 1.33. Use of reciprocity with dual networks.

We may sum up the foregoing results as: given a ladder network with source and load resistances R_s and R_L, another ladder network having an identical voltage transfer function (relating output voltage to source voltage) can be obtained by turning the dual network with respect to

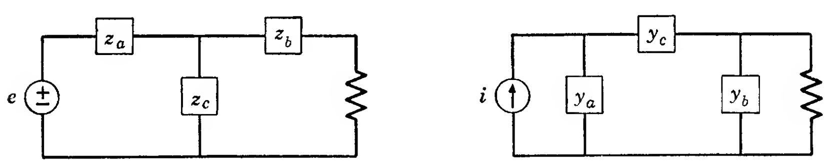


Fig. 1.34. T and pi networks.

$R_0{}^2 = R_s R_L$ end for end. (Note that the dual network is defined before turning the network around.)

Of special importance are the potential dual networks referred to as T and pi networks (alternative names are star and delta) shown in Fig. 1.34. These networks are of special importance because an equivalent T or pi at any given frequency may always be found for any arbitrary bilateral network, no matter how complex. Except at the given frequency, however, the equivalent T or pi may not always be physically

meaningful. Clearly, T and pi networks are simple examples of ladder networks. Another useful example of dual networks is the equivalence of voltage and current sources.

The ability to find a dual network is an extremely valuable tool in filter design. After a long and perhaps painful design procedure, the designer may come up with a filter. If the network has a dual, as is usually the case, the designer gets another network that accomplishes the same job as a free gift!

Perhaps the most enlightening way of finding a dual network makes use of the concepts presented in Sec. 1.9. Suppose we have a network driven with a voltage source for which we wish to find the dual with respect to R_0^2. The first thing to do is write the loop equations for the given network, which will be expressed in terms of the impedances $z_{ij}{}^m$. Then, multiply each and every term in each equation by G_0^2 (or divide by R_0^2). Then, interpret the self- and mutual terms as admittances $y_{ij}{}^n$ instead of impedances, interpret the loop currents as node voltages, interpret the driving voltages as driving currents (do not worry about proportionality constants for driving quantities), and finally, draw the corresponding node circuit. With respect to a driving voltage or current, the resulting dual networks will also have reciprocal input impedances.

1.11 Determinant manipulation for constant terminal conditions

Suppose that a set of equations describes a system having two pairs of terminals. The driving voltage or current occurs at node or loop 1 and the voltage at node k or current in loop k represents the output. It will be assumed that there is but a single driving current or voltage. Then, the transfer immittance is $\pm\Delta_{1k}/\Delta$. The input immittance is Δ_{11}/Δ, whereas the immittance looking into the network from the load end is Δ_{kk}/Δ. Suppose that we multiply all the elements of a row or all the elements of a column of Δ by a real positive number n. Then the determinant Δ becomes $n\Delta$. The various other determinants become $n\Delta_{11}$, $n\Delta_{1k}$, and $n\Delta_{kk}$ as long as the first or kth rows or columns are not modified. Then, the input, output, and transfer immittances, being the ratios of two determinants, are not changed because the factor n cancels. In other words, multiplying the jth row or column by n, where $j \neq 1$ and $j \neq k$, leaves the terminal conditions of the system unchanged. Clearly, in order for this manipulation to be done, the determinant Δ must have at least three rows and three columns.

Of course, after multiplying one row and column by a number n, a different row or column can be multiplied by a different number m with-

out changing the terminal conditions of the network as long as none of the rows or columns so modified is the first or the kth.

As an example, assume that we have the network of Fig. 1.35 and desire to obtain different networks having the same input impedance by multiplying the second row and/or column of the determinant of the

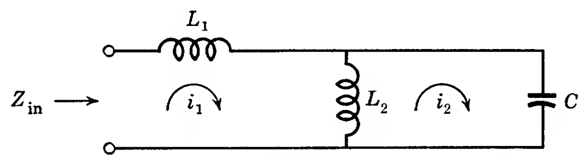


Fig. 1.35. Example of manipulation for constant input impedance.

network by n. If the equations of this network are written assuming loop currents as shown in Fig. 1.35, the determinant is

$$\begin{vmatrix} p(L_1 + L_2) & -pL_2 \\ -pL_2 & pL_2 + 1/pC \end{vmatrix} \tag{1.76}$$

Actually, a determinant is nothing but a number. An array of symbols as in eq. 1.76 is sometimes termed a "matrix." However, in order to avoid a discussion of matrix algebra, we shall continue to call such an array the determinant of the system.

If we multiply only the second row or only the second column by n, the determinant will not show symmetry about the main diagonal and hence will not represent a bilateral network. In order for it to represent a bilateral network, we must multiply both the second row *and* the second column by the *same* number n to give

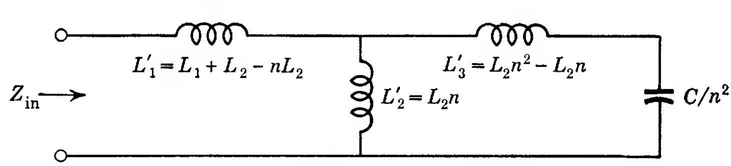


Fig. 1.36. The general equivalent circuit.

$$\begin{vmatrix} p(L_1 + L_2) & -pL_2n \\ -pL_2n & pL_2n^2 + 1/(pC/n^2) \end{vmatrix} \tag{1.77}$$

which represents the circuit of Fig. 1.36. In order that the inductances L_1' and L_3' of this figure be realizable, it is necessary that they both be positive and equal to or greater than zero. This puts bounds on the

number n as

$$1 \leq n \leq 1 + L_1/L_2 \tag{1.78}$$

If we let $n = 1$, the network is the same as that with which we started. If we let n be its maximum value, we get the quite different network of Fig. 1.37a. We could also use a transformer to give the circuit of Fig. 1.37b. The value of n required to make both series inductances of the

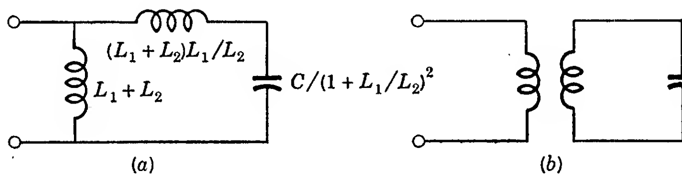


Fig. 1.37. Specific equivalent circuits.

equivalent circuit of Fig. 1.36 equal is found by equating L_1' and L_3' to be $(1 + L_1/L_2)^{1/2}$

The reader should observe that the number n can take on any value within its permissible range, a different network resulting from each choice for n, yet all these networks have identically the same input impedance. In other words, there are literally an infinite number of different networks that have the same input impedance.

As a second example, consider the four-node symmetric network of Fig. 1.38; the network has intentionally been made symmetric by splitting the series inductance L_2 into two equal inductances. In order to

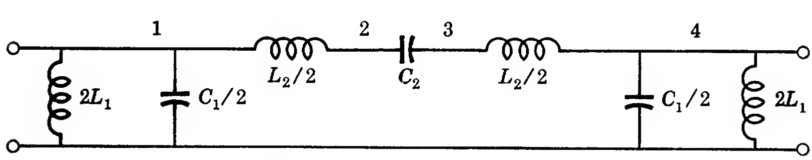


Fig. 1.38. A band-pass network.

maintain constant terminal conditions (at both ends), we can multiply the second and/or third rows or columns of the determinant by a number. If we want the resulting equivalent network to be bilateral, we must multiply the second row and the second column by the same number. Finally, if we want a physically symmetric network, the modified determinant must be symmetric with respect to both diagonals and we must symmetrically multiply the internal rows and columns. Let us assume that we wish to keep the network not only bilateral but also symmetric. Then, we must multiply the second and third rows and

columns by the same number n. The resulting determinant is

$$\begin{vmatrix} \left(\frac{pC_1}{2}+\frac{1}{p2L_1}+\frac{1}{\frac{pL_2}{2}}\right) & \left(\frac{-1}{\frac{pL_2}{2n}}\right) & (0) & (0) \\ \left(\frac{-1}{\frac{pL_2}{2n}}\right) & \left(pC_2n^2+\frac{1}{\frac{pL_2}{2n^2}}\right) & (-pC_2n^2) & (0) \\ (0) & (-pC_2n^2) & \left(pC_2n^2+\frac{1}{\frac{pL_2}{2n^2}}\right) & \left(\frac{-1}{\frac{pL_2}{2n}}\right) \\ (0) & (0) & \left(\frac{-1}{\frac{pL_2}{2n}}\right) & \left(\frac{pC_1}{2}+\frac{1}{p2L_1}+\frac{1}{\frac{pL_2}{2}}\right) \end{vmatrix} \tag{1.79}$$

This determinant typifies the network of Fig. 1.39, which is still physically symmetric. We see that the two pi circuits of inductance

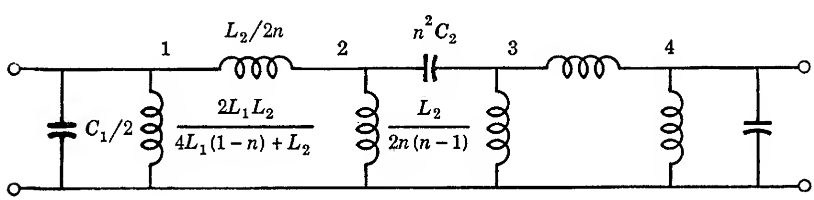


Fig. 1.39. The general band-pass equivalent network.

can be changed to practical transformers. For convenience and simplicity, let us make these pi networks in themselves symmetric. Then, we must have

$$\frac{2L_1L_2}{4L_1(1-n)+L_2} = \frac{L_2}{2n(n-1)} \tag{1.80}$$

from which

$$n = (1 + L_2/4L_1)^{1/2} \tag{1.81}$$

With this value of n, and using transformers in place of the pi equivalent, we get the circuit of Fig. 1.40, where $L = 2L_1$, and $M = 2L_1/$

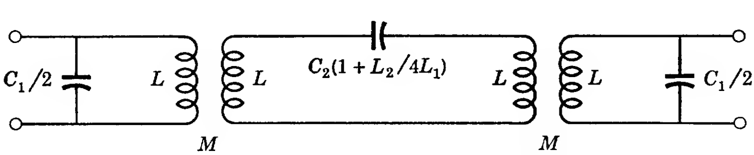


Fig. 1.40. The equivalent network with symmetric transformers.

$(1 + L_2/4L_1)^{1/2}$. Effectively, what we have done is to change the impedance level in the interior of the network without changing the terminal conditions of the network.

In order for determinant manipulation as described to work, it is necessary to have both mutual and self-immittances of the same type in the equation written for any one node or any one loop. For example,

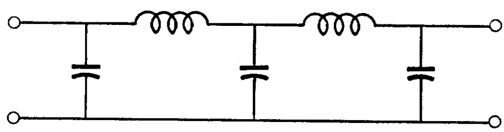


Fig. 1.41. A circuit not suited for determinant manipulation.

the reader can convince himself that the circuit of Fig. 1.41 cannot be changed without changing terminal conditions by using determinant manipulation as described. However, if an inductor should occur in shunt with any of the capacitors of Fig. 1.41, or if a capacitor should occur in series with an inductor, determinant manipulation will lead to equivalent circuits.

1.12 Determinant manipulation for changing impedance levels

If the elements of the first row and column of a determinant are multiplied by a number n where the driving source is in the first loop or node of the network, the input immittance becomes $\Delta_{11}/n^2\Delta$. Consequently, this manipulation changes the input immittance by the factor $1/n^2$. Of course, we could then go ahead and multiply any other row and column

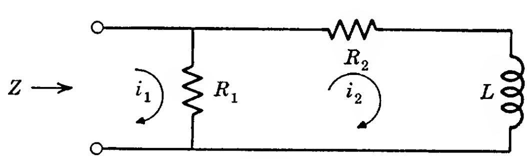


Fig. 1.42. Example for changing the input impedance level.

by a different number m and still have the input immittance changed by $1/n^2$.

As an example, consider the network having two loops shown in Fig. 1.42. The loop determinant of this network after manipulation as described is

$$\begin{vmatrix} n^2R_1 & -nmR_1 \\ -nmR_1 & (R_1 + R_2)m^2 + pLm^2 \end{vmatrix} \tag{1.82}$$

which represents the circuit of Fig. 1.43. In order that the two series resistances of this circuit be positive, it is necessary that

$$1 \leq n/m \leq 1 + R_2/R_1 \tag{1.83}$$

which permits the original input admittance $Y_{11}{}^m$ to be changed by *any* desired factor to $Y_{11}{}^m/n^2$ by suitably choosing n and m.

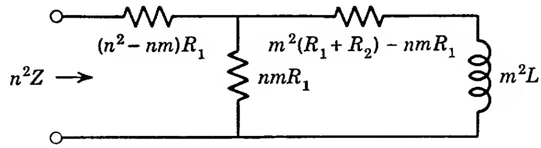


Fig. 1.43. The general equivalent circuit.

The particular choice $n = m$ is of special interest. In general (using a loop analysis), multiplying all rows and all columns of the determinant by the same number n results in a new network having the structural form of the original network but with all values of resistance and inductance multiplied by the factor n^2 and all values of capacitance divided by n^2. This is exemplified by the networks of Fig. 1.44.

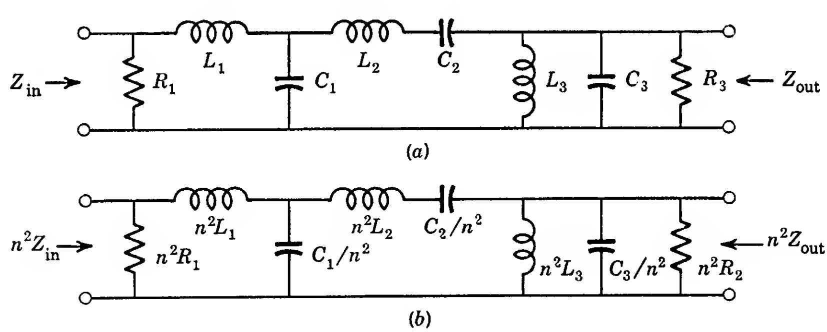


Fig. 1.44. Changing impedance level without structural changes.

Suppose that we have a network with two pairs of terminals and desire to change the impedance at the input but not at the output. Then, the first row and column can be multiplied by n. Of course, any other rows and columns except the one associated with the output node or loop can be multiplied by a different constant. By multiplying the first row and column by n, the input immittance is changed to $\Delta_{11}/n^2\Delta$, while the transfer immittance (assuming the source at the first node or loop and the load at the kth node or loop) is changed to $\Delta_{1k}/n\Delta$.

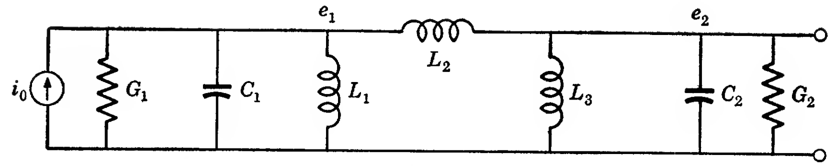


Fig. 1.45. Example for changing the transfer immittance.

As an example, consider the network of Fig. 1.45. The nodal determinant of this circuit with the first row and column multiplied by n is

$$\begin{vmatrix} \left(n^2G_1 + pn^2C_1 + \dfrac{1}{\dfrac{pL_1}{n^2}} + \dfrac{1}{\dfrac{pL_2}{n^2}}\right) & \left(\dfrac{-1}{\dfrac{pL_2}{n}}\right) \\ \left(\dfrac{-1}{\dfrac{pL_2}{n}}\right) & \left(G_2 + pC_2 + \dfrac{1}{pL_3} + \dfrac{1}{pL_2}\right) \end{vmatrix} \tag{1.84}$$

which is given by the circuit of Fig. 1.46.

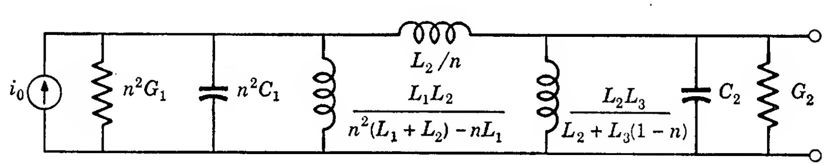


Fig. 1.46. Circuit with modified transfer immittance.

The value of this manipulation is that it finds a network that operates between source and load conductances n^2G_1 and G_2 from a network already designed to operate between conductances G_1 and G_2 without changing the behavior of the network with frequency and maintaining the same relative match between the source and load.

Problems

1. Prove the validity of the transformer equivalent circuit containing leakage inductance.
2. Prove the validity of the T equivalent to the transformer.
3. Prove the validity of the pi equivalent to the transformer.
4. Prove the validity of the general steady-state exchange of sources.
5. Prove the validity of the R, L, and C source exchanges using the differential equations and including initial conditions.

6. Starting with the infinite series definitions

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

show that

$$e^{\pm jx} = \cos x \pm j \sin x$$

7. From the results of Prob. 6, show that

$$\cos (x \pm y) = \cos x \cos y \mp \sin x \sin y$$

$$\sin (x \pm y) = \sin x \cos y \pm \cos x \sin y$$

8. From the results of Prob. 6, show that

$$\sin 2x = 2 \sin x \cos x$$

$$\cos 2x = \cos^2 x - \sin^2 x$$

9. From the results of Prob. 6, show that

$$\sin x = \frac{e^{jx} - e^{-jx}}{2j} \qquad \cos x = \frac{e^{jx} + e^{-jx}}{2}$$

10. From the results of Prob. 9, show that

$$\sin^2 x + \cos^2 x = 1$$

11. From the results of Prob. 6, derive the triple-angle equations

$$\sin 3x = 3 \sin x - 4 \sin^3 x$$

$$\cos 3x = 4 \cos^3 x - 3 \cos x$$

12. The hyperbolic functions can be defined by

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$$

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$$

Show that

$$\sinh x = \frac{e^x - e^{-x}}{2} = -j \sin jx$$

$$\cosh x = \frac{e^x + e^{-x}}{2} = \cos jx$$

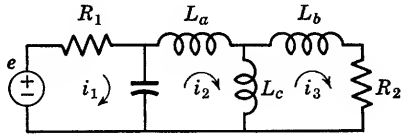


Fig. P.13.

13. For the circuit of Fig. P.13, find:
a. The three loop equations in p.
b. The determinant in terms of R, L, and C.
c. i_1/e and i_3/e as the ratio of two polynomials in p.

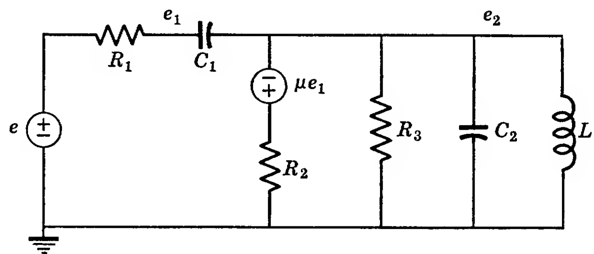


Fig. P.14.

14. For the circuit shown in Fig. P.14, find:
a. The two node equations in p.
b. The determinant in terms of R, L, and C.
c. e_1/e and e_2/e as the ratio of two polynomials in p.

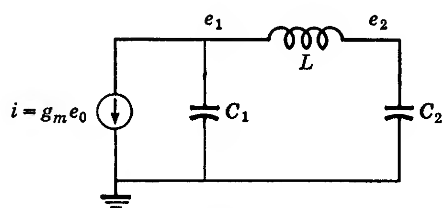


Fig. P.15.

15. Using node voltages, find e_1/e_0 and e_2/e_0 for the circuit shown in Fig. P.15 as the ratio of two polynomials in p.

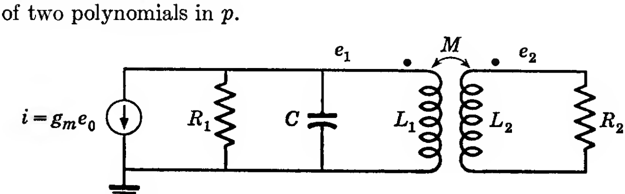


Fig. P.16.

16. Using loop currents and without using an equivalent circuit for the transformer, find e_2/e_0 for the circuit of Fig. P.16 as the ratio of two polynomials in p.
17. What is the answer to Prob. 16 if $C = 0$?
18. Suppose that $C = 0$ and $e_0 = e_a + ke_2$ in Prob. 16, where k is a constant and e_a is the driving voltage. What is e_2/e_a as the ratio of two polynomials in p?

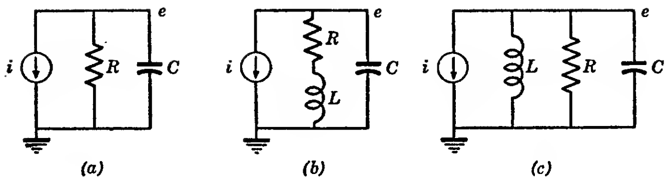


Fig. P.19.

19. Find $e/i = Z$ for the circuits of Fig. P.19 as ratios of two polynomials in p.

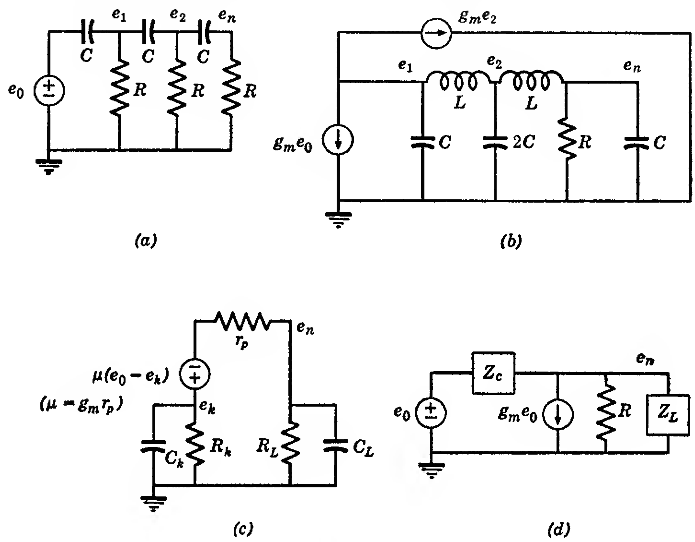


Fig. P.20.

20. Write the node equations (in terms of p) for the circuits of Fig. P.20. Solve for e_n/e_0 in all cases.

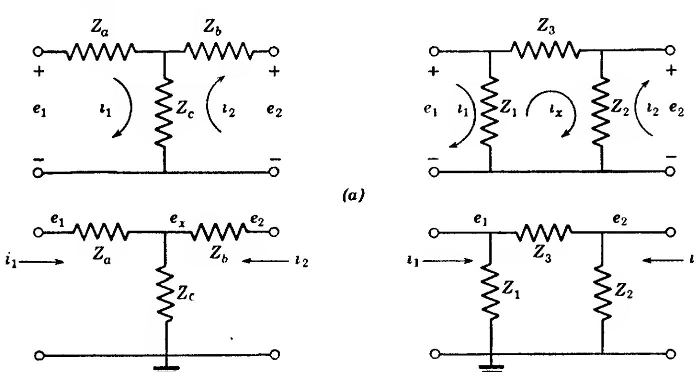


Fig. P.21.

21. Write the loop equations for the circuits of Fig. P.21*a*. Eliminate i_x from the equations and find Z_a, Z_b, and Z_c in terms of Z_1, Z_2, and Z_3 so that the terminal conditions of the two networks are identical. Similarly, find Z_1, Z_2, and Z_3 in terms of Z_a, Z_b, and Z_c by solving the circuits of Fig. P.21*b* using node voltages. The results of this problem give the transformations between T and pi networks.

22. Prove the theorem of maximum power transfer.

23. A spring-mass system with viscous damping has the equation

$$m\frac{d^2x}{dt^2} + f\frac{dx}{dt} + kx = F$$

where x = distance, m = mass, k = spring constant, f = viscous damping constant, and F = force. Writing v = velocity for dx/dt, show two analog electric circuits and identify the analogous variables.

24. A set of equations is

$$(R_1 + Lp + S_1/p)i_1 - Lpi_2 = e$$
$$-Lpi_1 + (R_2 + Lp + S_2/p)i_2 = 0$$

Draw the circuit represented by these equations.

25. Find the circuits whose input impedances are reciprocal with respect to R_0 to the circuits of Fig. P.25.

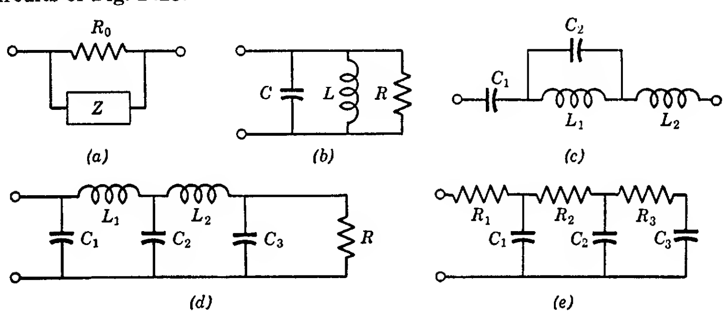


Fig. P.25.

26. Obtain the duals (with respect to R_0) of the networks of Fig. P.26.

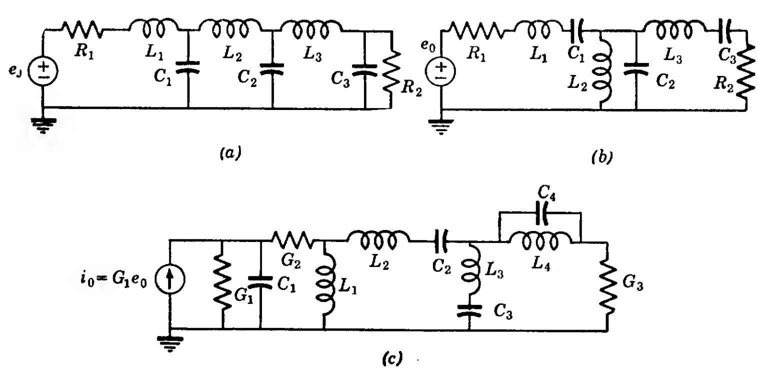


Fig. P.26.

27. The circuit shown in Fig. P.27*a* containing an ideal transformer can be converted to a practical transformer or possibly to a practical T or pi network of inductances. The first step is to "refer" quantities as shown in Fig. P.27*b*. Then, either a

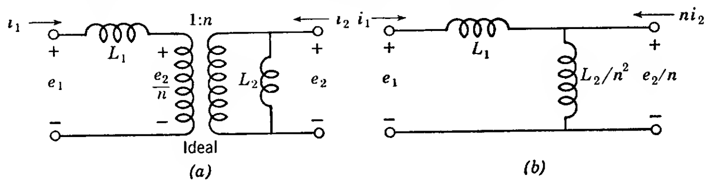


Fig. P.27.

loop or a node analysis can be used to find the equivalent T or pi network. Determine the T and pi networks of inductances for the circuit. Determine the limits on n in order that the equivalent T or pi be physical. These limits will depend on the ratio of L_1 to L_2.

28. Repeat Prob. 27 for the circuit shown in Fig. P.28.

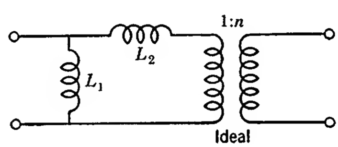


Fig. P.28.

29. The circuit containing an ideal transformer shown in Fig. P.29*a* can sometimes be converted to a T or pi network of capacitors. Find the equivalent T and pi networks using a method like that of Prob. 27. Determine the limits on n in order that the equivalent circuit have nonnegative elements.

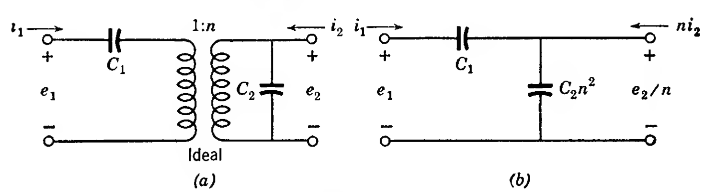


Fig. P.29.

30. Repeat Prob. 29 for the circuit shown in Fig. P.30.

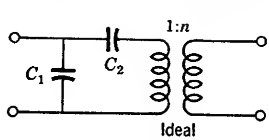


Fig. P.30.

31. Consider the filter described by Fig. P.31. The ratio $|e_2/e_0|$ is to have the shape as a function of frequency as indicated. What is this ratio at $\omega = 0$? Is it

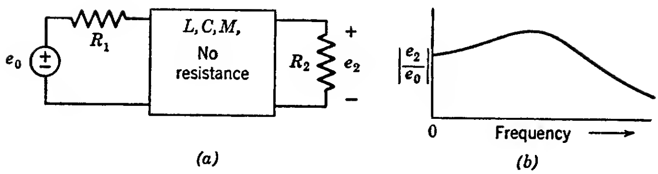


Fig. P.31.

possible to use a practical transformer if $R_1 \neq R_2$ in order to have the load and source resistances matched for maximum power transfer at $\omega = 0$? Discuss. What is the maximum possible value of $|e_2/e_0|$ at a finite frequency?

32. Given the network of Fig. P.32:

a. Show the equivalent network having the same Z_{in}. (Express element values in terms of the multiplying constant.)

b. Show the equivalent network with an impedance Z_{in}/k^2 obtained by multiplying only the first row and column by k. What are the limits on k?

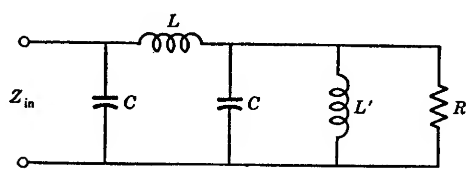


Fig. P.32.

c. Show the equivalent circuit with impedance Z_{in}/k^2 obtained by multiplying both rows and both columns by k. What is k if $R = 1\ \Omega$?

d. Show the equivalent circuit with impedance Z_{in}/k^2 obtained by multiplying the first row and column by k and the second row and column by n. In this case, what must be the relation between n and k in order that the equivalent circuit have positive element values?

e. Can equivalents having the impedance Z_{in} be found using determinant manipulation of L' does not appear in the circuit?

33. Given the network of Fig. P.33:

a. Multiply the first row and first column by k. How does this affect the transfer impedance e_2/i_0? What bounds are placed on k? Show the new element values.

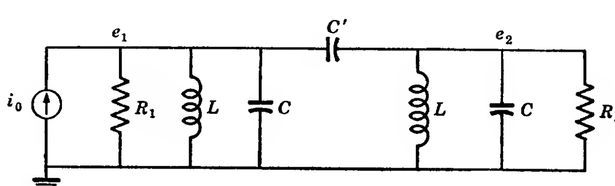


Fig. P.33.

b. Same as part *a* but multiply the second row and column by n instead of the first row and column by k.

c. Multiply the first row and column by k and the second row and column by n. Repeat part *a*.

34. Normalize the circuits of Fig. P.34 to a resistance $R_x = 1\ \Omega$ without changing the structural form of the networks. After normalization to 1 Ω, go through the reverse procedure to make R_x ten times that shown in the figure.

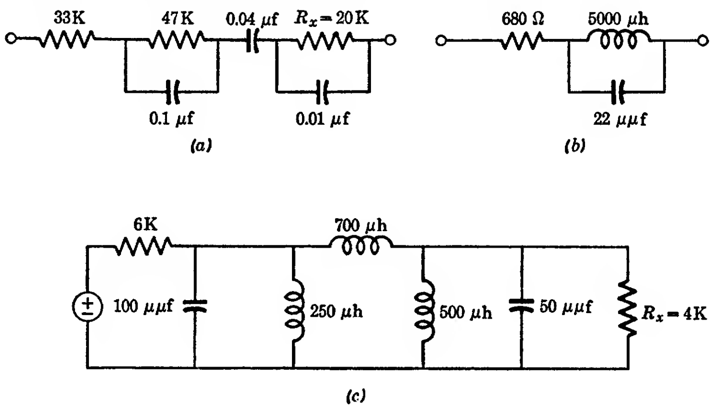


Fig. P.34.

35. Using determinants, prove the theorem of reciprocity.

36. In the theory of image-matched filters, certain special filter structures play important roles. Find e_2/i_2, e_1/i_1, e_1/i_2, and e_2/i_1 for each of the networks shown in Fig. P.36. The labor of finding the required functions can be reduced by using the symmetric property of most of the networks.

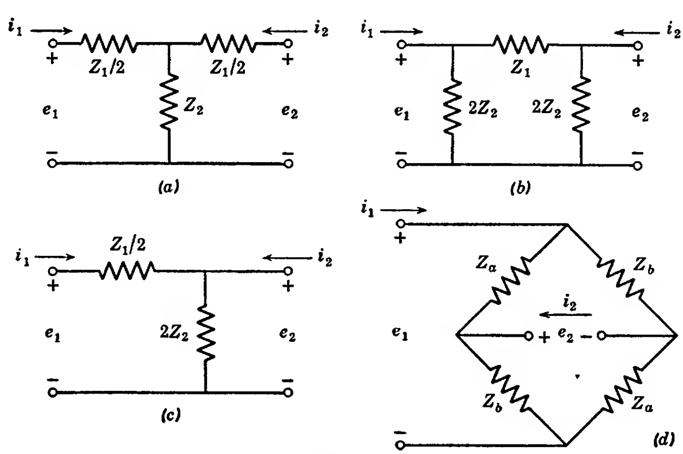


Fig. P.36.

2

The Poles and Zeros of Functions

The theory of functions of a complex variable is an important branch of mathematics which treats variables that are not simply real positive or negative quantities but have both real and imaginary parts. From high-school studies of algebra, the student gains a facility in handling real functions of a real variable. In elementary college courses in network analysis, the electrical engineer gains a facility in treating complex functions of the purely imaginary variable $j\omega$. The complex variable is simply a combination of two variables, one real and one complex.

In this chapter, we wish to study some of the aspects of functions of a complex variable and rational functions in particular. Since this book does not profess to be a text on mathematics, much of what will be said will be simply stated with perhaps a justification that is not too rigorous. We shall have nothing to say about the integration of functions of a complex variable, which is perhaps the most fruitful mathematical part of the theory. Rather, we shall restrict ourselves to those parts of the theory that have the most widespread practical value in furnishing a thorough pictorial understanding of network functions.

2.1 General functions

Consider some arbitrary function of the variable x, $F(x)$. As x varies in magnitude, $F(x)$ goes through gyrations of one sort or another. For certain values of x, x_i, the function $F(x)$ may become zero. For example, $F(x) = x - 1$ becomes zero at $x = 1$. These particular values of x are called the "zeros" of $F(x)$.

For some other values of x, x_k, the function $F(x)$ may go to infinity. For example, $1/(1 + x)$ goes to infinity at $x = -1$. Such specific values of x are called the "singularities" of $F(x)$. If $F(x)$ behaves as $1/(x - x_k)^q$ in the immediate vicinity of x_k, where q is a positive integer, the singularity at $x = x_k$ is termed a "pole" of order q of $F(x)$. A similar ter-

minology is used to define a zero of order q of $F(x)$ when $F(x)$ behaves as $(x - x_i)^q$ in the vicinity of the zero x_i.

We are all familiar with poles and zeros (but perhaps not with the terminology) when the p-z (shorthand for poles and zeros or pole-zero) of the function $F(x)$ are ordinary numbers. For example, the function $\sin x$ has zeros at $x = \pm n\pi$, where n is any integer but has no poles. The function $\csc x = 1/\sin x$ has poles at $x = \pm n\pi$, where n is any integer but has no zeros. The function $\tan x$ has both p-z, the poles being at $x = \pm n\pi/2$ for n odd, and the zeros at $\pm n\pi/2$ for n even.

Frequently, a function $F(x)$ has few or no p-z for ordinary values of x. By assuming that the variable x is really a *complex variable* having both a real and an imaginary part, many more p-z may be obtained.

Let the complex variable be represented by $p = \sigma + j\omega$ where σ and ω are the real and imaginary parts of p respectively. A function of a complex variable can then be written

$$F(p) = F(\sigma + j\omega) = F_1(\sigma, \omega) + jF_2(\sigma, \omega) \tag{2.1}$$

where F_1 and F_2 are the real and imaginary parts of F respectively.

The function F_1 is the equation of a surface; it is the height of a surface above a two-dimensional plane defined by orthogonal axes along which σ and ω are measured. Similarly, F_2 is the equation of another surface. Consequently, a function of a complex variable $p = \sigma + j\omega$ can be interpreted as two surfaces above a (σ, ω) plane, the height of one surface giving the real part of the function, and the height of the other giving the imaginary part of the function. The (σ, ω) plane is called the "p plane." The horizontal axis of this plane is called the real or σ axis, and the vertical axis is called the imaginary or $j\omega$ axis (or sometimes simply the j axis).

Expressing the real and imaginary parts of a function of a complex variable with the heights of surfaces above the p plane is not the only way that the function can be represented. It may also be expressed in terms of a magnitude and a phase angle as

$$\begin{aligned} |F(p)| &= [F_1(\sigma, \omega)^2 + F_2(\sigma, \omega)^2]^{1/2} \\ \arg F(p) &= \tan^{-1} F_2/F_1 \end{aligned} \tag{2.2}$$

where "arg" stands for argument or angle. Thus, $|F(p)|$ may be described as one surface above the p plane and $\arg F(p)$ as another. This representation is the one we shall employ.

Ordinary functions of a single real variable such as sine, cosine, tangent, exponential, hyperbolic functions, or polynomials, and products, quotients, and sums of these functions can be converted to functions of a complex variable through the simple expedient of replacing the real

variable x with a complex variable $p = \sigma + j\omega$. For example

$$\sin x \rightarrow \sin p = \sin (\sigma + j\omega) \tag{2.3}$$

All of the functions of a complex variable obtained in this manner are called analytic functions because they have derivatives independent of direction on the p plane which are bounded in magnitude everywhere except at the exact positions of the singularities of the function. In addition to a constant multiplier, *analytic functions are completely defined by the values of the complex variable at which the function becomes zero and infinite.*

The values of p at which the function becomes zero are the zeros of the function z_i, and the values of p at which the function becomes infinite are the singularities (usually the poles) of the function p_i. For multiple-order poles or zeros, several may exist at the same point on the p plane.

It must be pointed out that the definition we have used to describe analytic functions by no means includes all analytic functions. However, it does include all those having a relationship to physical systems. For reasons that will become apparent later, we shall call this class of analytic functions "conjugate analytic."

The values of p at which the function $F(p)$ becomes zero or infinite may be plotted on the p plane. An example of this is shown in Fig. 2.1 where a zero is designated by a circle and a pole by a cross. A pole

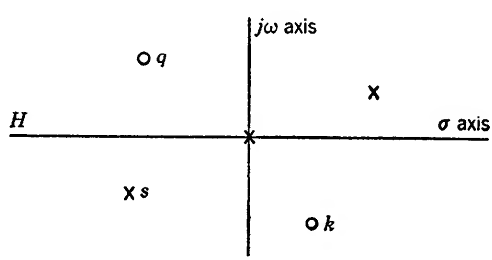


Fig. 2.1. Notation for the p plane.

(or zero) of order q is designated by a circle or cross with the number q close by. No number is specified if the order is unity. The constant multiplier H can be put in evidence somewhere on the p plane. *It is most important to realize that a sketch such as that of Fig. 2.1 is a complete representation of the analytic function.*

When the variable p is near a pole, the magnitude of the function is large, and when the variable p is near a zero, the magnitude of the

function is small. Between p-z, there is a continuous transition in the magnitude of the function.

It is convenient to separate the p plane into three distinct regions. The total area to the left of the imaginary axis is the most important one and is called the "left half-plane." The total area to the right of the imaginary axis is called the "right half-plane." The third region, which is infinitesimally small in area, is the imaginary axis itself.

A pole position p_1 is designated by $p_1 = \sigma_1 + j\omega_1$, a two-dimensional vector which will henceforth be referred to as a phasor. Similarly, the

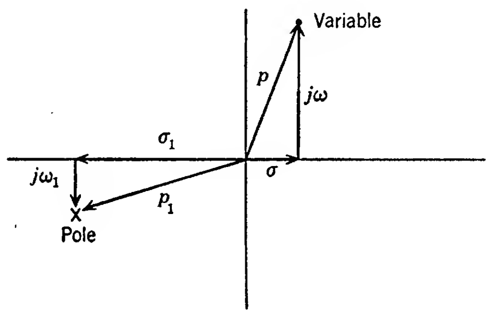


Fig. 2.2. Phasors for a pole and the variable.

variable $p = \sigma + j\omega$ is a phasor. This can be seen more clearly by referring to Fig. 2.2. *The pole p_1 is fixed and is characteristic of the function, whereas the variable p can occur anywhere on the p plane, depending upon where an evaluation of the function is desired.*

It should be observed that some functions may have poles of infinite order which may have positions that are either finite or infinite. A function is said to have an essential singularity at the position of an infinity of poles. For example, the exponential e^x has an essential singularity at $x = \infty$. The functions of principal interest in network theory do not have essential singularities.

Let a general analytic function have zeros at $p = z_1, z_2, z_3, \cdots$ with orders $e_1, e_2, e_3, \cdots$ respectively. Let it also have poles at $p = p_1, p_2, p_3, \cdots$ with orders $f_1, f_2, f_3, \cdots$ respectively. There may be an infinite number of p-z. Such an analytic function may *always* be written

$$F(p) = H\frac{(p - z_1)^{e_1}(p - z_2)^{e_2}(p - z_3)^{e_3}\cdots}{(p - p_1)^{f_1}(p - p_2)^{f_2}(p - p_3)^{f_3}\cdots} \tag{2.4}$$

which is a product of factors divided by a similar product of factors. (An arbitrary constant can also be added to eq. 2.4.)

An important but often overlooked fact is that should a pole position be equal to a zero position the relevant factors in eq. 2.4 cancel; that is, a pole cancels a zero if it occurs at the same point on the p plane. It will be assumed in our discussion of functions in this chapter that such identical factors have been cancelled.

Let us examine a typical factor in eq. 2.4, for example, $p - z_1$. By studying Fig. 2.3, it can be seen that $p - z_1$ represents an ordinary phasor starting at the zero and terminating at the point at which it is

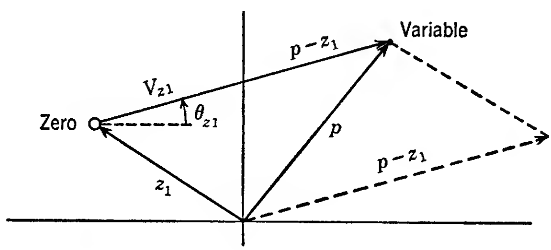


Fig. 2.3. Phasor interpretation of a factor.

desired to evaluate the factor $p - z_1$. Let this phasor be called $V_{z1}\underline{/\theta_{z1}}$, where $\underline{/\theta}$ signifies the angle of the phasor which is measured in the counterclockwise sense from the direction of the positive real axis (which follows convention). Since all other factors can be treated similarly, it follows that eq. 2.4 can be expressed as

$$F(p) = H\frac{(V_{z1}\underline{/\theta_{z1}})^{e_1}(V_{z2}\underline{/\theta_{z2}})^{e_2}\cdots}{(V_{p1}\underline{/\theta_{p1}})^{f_1}(V_{p2}\underline{/\theta_{p2}})^{f_2}\cdots} \tag{2.5}$$

in which the subscript zi represents the ith zero and the subscript pi represents the ith pole.

Finally, eq. 2.5 can be put in a most significant and convenient form by using the concepts of ordinary complex algebra in order to express $F(p)$ as a magnitude and a phase angle as

$$|F(p)| = H\frac{(V_{z1})^{e_1}(V_{z2})^{e_2}\cdots}{(V_{p1})^{f_1}(V_{p2})^{f_2}\cdots} \tag{2.6}$$

$$\arg F(p) = (e_1\theta_{z1} + e_2\theta_{z2} + \cdots) - (f_1\theta_{p1} + f_2\theta_{p2} + \cdots)$$

The magnitude of the function is given by the product of all the phasor lengths from all the zeros divided by the product of all the phasor lengths from all the poles, where all the phasors terminate at the point p at which an evaluation of the function is desired. The phase angle of the function

is the sum of the angles of all the phasors associated with the zeros diminished by the sum of the angles of all the phasors associated with the poles.

If desired, the two surfaces above the p plane representing the magnitude and the phase angle can be constructed from the data provided by eqs. 2.6. However, we shall not require evaluations of $F(p)$ over the entire p plane but only along one of the two axes: the imaginary axis, to be specific. Thus, we shall be interested only in cross sections

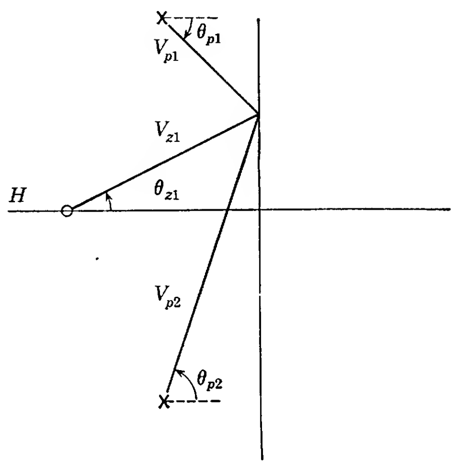


Fig. 2.4. Evaluation along the imaginary axis.

of the surfaces, which yield curves of the magnitude and phase angle of $F(p)$ for values of p along the $j\omega$ axis.

We shall usually represent functions of interest to us with p-plane plots of the poles and zeros which, when viewed in the light of an equation such as eq. 2.6, give a qualitative understanding of the variation of the magnitude and phase of the function of the complex variable as the variable is allowed to roam over the p plane. As an example, consider the relatively simple function described by the p-z of Fig. 2.4. Assume that it is desired to evaluate the function along the $j\omega$ axis as indicated in Fig. 2.4. Then

$$\begin{aligned} F(p) &= H \frac{V_{z1}\underline{/\theta_{z1}}}{V_{p1}\underline{/\theta_{p1}}V_{p2}\underline{/\theta_{p2}}} \\ &= H \frac{V_{z1}}{V_{p1}V_{p2}} \underline{/\theta_{z1} - (\theta_{p1} + \theta_{p2})} \end{aligned} \tag{2.7}$$

The sketches of Fig. 2.5 for the magnitude and phase angle can be obtained practically by inspection. They may be obtained with some accuracy by using a protractor (or better still, a drafting machine) and a pair of dividers. The length and angle of each phasor can be found separately and then combined to give the evaluation of the entire function.

Consider a purely real function of a purely real variable $F(x)$. Obviously, $F(x)$ can be either positive or negative but can never have an

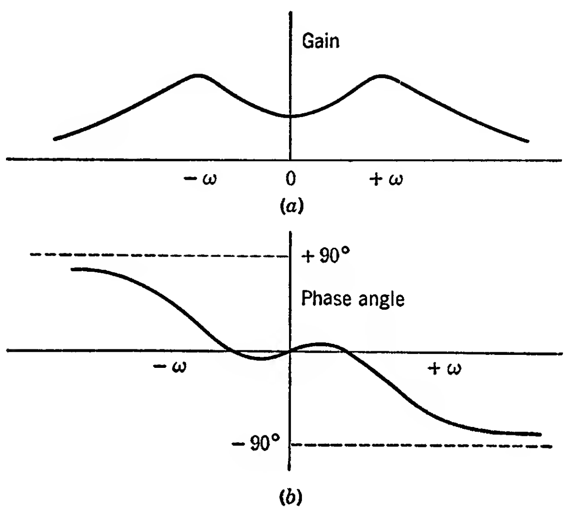


Fig. 2.5. Gain and phase functions.

imaginary part. Let this function be converted to one of a complex variable by substituting $p = \sigma + j\omega$ for x to give the function $F(p)$, which is now characterized by the positions of its p-z and which has been defined as a conjugate-analytic function. If the function of p is evaluated along the real axis of the p plane, we have $F(\sigma + j0) = F(\sigma)$, which is the original real function of a real variable. Thus, *a function of a real variable is merely a special case of the more general function of a complex variable,* and functions of real variables as well as functions of complex variables can be described by p-z on the complex-variable plane.

Since $F(x) = F(\sigma)$ can never be complex but must always be either positive or negative real, the phasors from the p-z to the real axis of the p plane must have phase angles that somehow cancel. This is possible in general only if each pole or zero at $p = \sigma_i + j\omega_i$ is accompanied by its complex-conjugate pole or zero at $p = \sigma_i - j\omega_i$. Thus, *the p-z of conjugate-analytic functions, if complex, must occur in complex-conjugate*

pairs. This definition clearly distinguishes conjugate-analytic functions from all other kinds of analytic functions.

As an example, consider the real function $F(x) = x^2 + 1$, which has no poles or zeros for purely real values of x. Converting this, we get the complex function $F(p) = p^2 + 1 = (p - j)(p + j)$, which has two zeros located at z_1, $z_1^* = \pm j$, which are purely imaginary and are conjugates. Along the real axis, $p = \sigma$ and $F(\sigma) = \sigma^2 + 1$, which is the

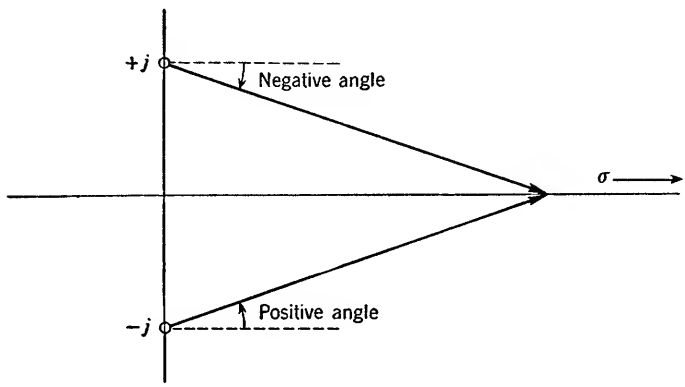


Fig. 2.6. Evaluation along the real axis.

original real function of a real variable. On the p plane, the evaluation is shown in Fig. 2.6. The length of each phasor is $(1 + \sigma^2)^{1/2}$; hence, the product of the two is $1 + \sigma^2$. The angle is zero because the negative and positive angles from the two zeros cancel.

2.2 Polynomial functions

Whenever the number of p-z of a function is finite, the factored expression for the function consists of a finite number of factors in both numerator and denominator with each factor raised to a finite power. Let there be such a function having m zeros and n poles, some of which may be of multiple order. Then, the function can be described by

$$F(p) = H \frac{(p - z_1)(p - z_2) \cdots (p - z_m)}{(p - p_1)(p - p_2) \cdots (p - p_n)} \tag{2.8}$$

The factors in eq. 2.8 can be multiplied out to give an expression consisting of the ratio of two polynomials in p as

$$F(p) = H \frac{p^m + a_{m-1}p^{m-1} + a_{m-2}p^{m-2} + \cdots + a_1 p + a_0}{p^n + b_{n-1}p^{n-1} + b_{n-2}p^{n-2} + \cdots + b_1 p + b_0} \tag{2.9}$$

The polynomial in the numerator is of degree m and that in the denominator is of degree n. The a's and b's are constant coefficients. The p-z, if not purely real, must occur in complex-conjugate pairs since it is assumed that $F(p)$ is the counterpart of a real function of a real variable; that is, it is a conjugate-analytic function. Since m and n may be arbitrarily large, the function of eq. 2.9 may approximate *any* conjugate-analytic function with arbitrary accuracy. In fact, a function that is the ratio of two polynomials is the most powerful type of function capable of approximating an arbitrary real function of a real variable. This property causes functions of the type given by eq. 2.9 to be about as general as is possible. (See any text on complex-variable theory which discusses Taylor and Laurent expansions.)

Let us investigate the relations between the a's and b's of eq. 2.9 and the p-z positions $z_1, z_2, \cdots$ and $p_1, p_2, \cdots$. Multiplying out eq. 2.8 and equating the results to eq. 2.9, we observe that

$$\begin{aligned} -a_{m-1} &= \text{Sum of the zero positions} \\ -b_{m-1} &= \text{Sum of the pole positions} \end{aligned} \qquad (2.10)$$

It can be observed that since the p-z must occur as complex-conjugate pairs, the sums of eqs. 2.10 are the sums of only the real parts of the p-z positions.

Similarly, a_{m-2} can be seen to be the sum of the products of the zero positions taken two at a time, and so forth. Finally

$$\begin{aligned} \pm a_0 &= \text{Product of all the zero positions} \\ \pm b_0 &= \text{Product of all the pole positions} \end{aligned} \qquad (2.11)$$

where the sign depends on whether n and m are even or odd.

Because all p-z occur in complex-conjugate pairs if complex, it can be seen that *the coefficients a_k and b_k are purely real.* The reader can easily prove this for himself.

A function such as that of eq. 2.9 is called a "rational meromorphic," or more simply, a "rational" fraction. Such functions are always composed of the quotient of two polynomials in the variable p.

That eq. 2.9 can always be written in factored form follows from the fundamental law of algebra which states that every polynomial of degree higher than the zeroth has at least one zero; that is

$$(p^n + a_{n-1}p^{n-1} + \cdots + a_0) = (p - z_1)(p^{n-1} + c_{n-2}p^{n-2} + \cdots + c_0) \qquad (2.12)$$

Applying the same theorem to the polynomial remaining on the right of eq. 2.12 results in the algebraic theorem: A polynomial of degree n has n zeros.

Another fundamental law of algebra states that the complex zeros of a polynomial having purely real coefficients must occur in complex-conjugate pairs. In other words, polynomials having real coefficients are conjugate-analytic functions.

These two laws of algebra are not really necessary to the discussion here since we showed earlier that they must hold true for conjugate-analytic functions. However, had we begun by studying unfactored polynomials having real coefficients rather than analytic functions, these two laws of algebra would prove the existence of zeros and that the zeros, if complex, must occur in complex-conjugate pairs.

A zero on the negative real axis at $z_1 = -\delta$ in a factored polynomial occurs as a factor $p + \delta$. A pair of complex-conjugate zeros in the left half-plane with positions $z_2, z_2^* = -\alpha \pm j\beta$ occurs as a factor $p^2 + 2\alpha p + (\alpha^2 + \beta^2)$. Thus it can be seen that when all the zeros of a polynomial lie in the left half-plane, the polynomial resulting after multiplying out all the factors relating to the zero positions must contain all powers of p and all coefficients must be positive and real. If some of the powers of p of a polynomial are missing, or if any of the coefficients are negative, the polynomial will have one or more zeros on the imaginary axis or within the right half-plane. It is important to observe that negative coefficients or missing powers of p constitute a sufficient but not necessary condition that zeros lie in the right half-plane. *A polynomial having zeros in the left half-plane only is called a "Hurwitz" polynomial.*

A brief word at this point may save the reader some confusion. A rational fraction may have both poles and zeros. When we speak of the polynomial in either the numerator or the denominator of a rational fraction, we speak of a function containing only zeros. If it is the polynomial in the denominator of the rational fraction which we happen to be studying, then the zeros of the polynomial are the poles of the rational fraction. Again, a polynomial by itself has only zeros.

Consider a rather special polynomial having zeros only on the $j\omega$ axis. Each pair of conjugate zeros will be described by a factor $(p + j\omega_1)(p - j\omega_1) = p^2 + \omega_1^2$. A zero at the origin is described by the factor $(p - 0)$. Thus, *a polynomial having zeros only on the imaginary axis of the p plane will have only even powers of p or only odd powers of p and all the coefficients will be positive.* Such polynomials can be termed "reactance" polynomials.

If the even and odd powers of p of a Hurwitz polynomial are grouped together as indicated in eqs. 2.13, two reactance polynomials will result.

$$\begin{aligned} F(p) &= p^7 + a_6p^6 + a_5p^5 + a_4p^4 + a_3p^3 + a_2p^2 + a_1p + a_0 \\ &= \text{Ev } F(p) + \text{Od } F(p) \qquad (2.13) \\ &= (a_6p^6 + a_4p^4 + a_2p^2 + a_0) + p(p^6 + a_5p^4 + a_3p^2 + a_1) \end{aligned}$$

Operators Ev and Od stand for the operations of taking the even and odd powers, respectively, of the polynomial on which they operate.

Another kind of polynomial which has only even powers of p but

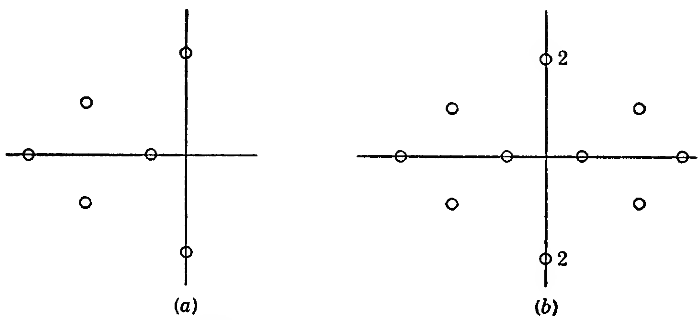


Fig. 2.7. Quadrantal symmetry.

which may have both positive and negative coefficients is one displaying "quadrantal symmetry." Let pairs of complex-conjugate zeros in the left half-plane and real zeros on the negative real axis be mirrored about the $j\omega$ axis of the p plane. A zero on the $j\omega$ axis is mirrored by another zero at the same point. The total of zeros displays perfect symmetry with respect to the imaginary axis of the p plane. A plot with quadrantal symmetry is exemplified by Fig. 2.7*b*. The reader should study this figure in order to satisfy himself that *a function having quadrantal symmetry has a phase angle that is zero (or 180) degrees along the entire $j\omega$ axis of the p plane.*

Let us now examine the significance of quadrantal symmetry by studying one pair of complex-conjugate zeros. Since such a pair of zeros represents one factor in an expression that may have many similar factors, the extension to many zeros is trivial. Assume that a pair of zeros exists in the left half-plane at $z_1, z_1^* = -\alpha \pm j\beta$. Then, the factor relating to these zeros is

$$F(p) = p^2 + 2\alpha p + (\alpha^2 + \beta^2) \qquad (2.14)$$

The mirror images of these zeros in the right half-plane are at z_2, $z_2^* = +\alpha \pm j\beta$. The only difference between the positions of z_1, z_1^* and z_2, z_2^* is that the real parts are negatives of one another; hence, z_2, z_2^* are called the "negatives" of z_1, z_1^*. The polynomial relating to the zeros z_2, z_2^* is

$$F_{\text{neg}}(p) = p^2 - 2\alpha p + (\alpha^2 + \beta^2) = F(-p) \tag{2.15}$$

where it is to be noted that, if a group of zeros is defined by $F(p)$, the negatives of these zeros is described by $F(-p)$.

The function representing all four zeros is

$$\begin{aligned} F(p)F(-p) &= (p^2 + 2\alpha p + \alpha^2 + \beta^2)(p^2 - 2\alpha p + \alpha^2 + \beta^2) \\ &= p^4 + [2(\alpha^2 + \beta^2) - 4\alpha^2]p^2 + (\alpha^2 + \beta^2)^2 \end{aligned} \tag{2.16}$$

which has only even powers of p.

Let us evaluate $F(p)$ at $p = j\omega$, that is, along the imaginary axis of the p plane

$$\begin{aligned} F(j\omega) &= -\omega^2 + 2\alpha j\omega + \alpha^2 + \beta^2 \\ F(-j\omega) &= -\omega^2 - 2\alpha j\omega + \alpha^2 + \beta^2 \end{aligned} \tag{2.17}$$

But since $F^*(j\omega) = F(-j\omega)$, we obtain

$$[F(p)F(-p)]_{p=j\omega} = F(j\omega)F^*(j\omega) = |F(j\omega)|^2 \tag{2.18}$$

We conclude that, along the $j\omega$ axis, the square of the magnitude of the function $F(j\omega)$ is given by a function containing the p-z of $F(p)$ plus their negatives.

2.3 Relationship to networks

In the steady state, networks are characterized by a collection of input and transfer immittances that are functions of $p = j\omega$. In fact, it is not difficult to see that these immittances will have the form of the ratio of two polynomials in $p = j\omega$ with real coefficients. That the coefficients are real follows from the fact that they are made up of products, sums, and differences of the element values R, L, C, and M (and g_m, r_p, and μ when vacuum tubes are involved).

The immittances are evaluated in the steady state by setting p equal to $j\omega$ and using complex algebra to obtain a solution, which amounts to *evaluating a rational function of a complex variable along the positive imaginary axis of the p plane.* Hence, we can interpret the differential operator p in network equations to be the complex variable p. Then input and transfer immittances can be evaluated along the positive $j\omega$ axis of the p plane to give the steady-state characteristics.

In order to convert equations written in terms of p or $j\omega$, it is only necessary to substitute $p = \sigma + j\omega$ for the differential operator p or for $j\omega$. The resulting equations can be termed the "p equations." Similar notation was used in Chapter 1 as a matter of convenience and definition.

Since network functions have the form of rational fractions with real coefficients, their p-z must either be purely real or occur in complex-conjugate pairs; that is, they must be conjugate-analytic functions. This symmetric distribution of p-z means that along the imaginary axis (both positive and negative) *the magnitude of an immittance function must be an even function of frequency, whereas its phase angle must be an odd function of frequency.* This is exemplified by the curves of Fig. 2.5.

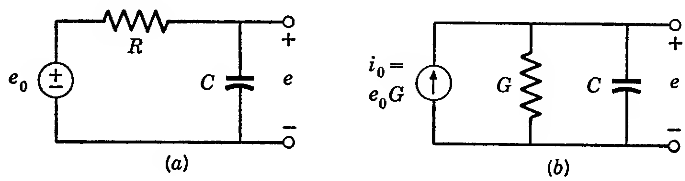


Fig. 2.8. The *R-C* circuit.

Let us see how the p-z concept applies to one of the simplest circuits that varies with frequency, that is, to the *R-C* circuit shown in Fig. 2.8. Let it be assumed that the transfer function e/e_0 is desired. With conventional complex algebra, we obtain

$$\frac{e}{e_0} = \frac{1/j\omega C}{R + 1/j\omega C} = \frac{1/RC}{j\omega + 1/RC} \tag{2.19}$$

or

$$\frac{e}{e_0} = \frac{1/RC}{[(1/RC)^2 + \omega^2]^{1/2}} \underline{/-\tan^{-1}\omega RC} \tag{2.20}$$

Let us plot the p-z of this circuit. To do this, p is substituted for $j\omega$ in eq. 2.19 to get

$$\frac{e}{e_0} = \frac{1/RC}{p + 1/RC} \tag{2.21}$$

which has but a single pole at $p = -1/RC$. The constant multiplier is $1/RC$. Thus, the p-plane plot of Fig. 2.9 is a *complete representation* of the transfer function.

From Fig. 2.9 rather than from preceding equations

$$\frac{e}{e_0} = \frac{H}{p - p_1} = \frac{1/RC}{V} = \frac{1/RC}{|V| \underline{/\theta}} = \frac{1/RC}{|V|} \underline{/-\theta} \tag{2.22}$$

Clearly, θ is the angle whose tangent is b/a in Fig. 2.9. Since $b = \omega$ and $a = 1/RC$, $\theta = \tan^{-1} \omega RC$, which is in agreement with eq. 2.20. The magnitude of V is also found to be in agreement with eq. 2.20.

An inspection of Fig. 2.9 tells at a glance how the magnitude of e/e_0 decreases with frequency and how the phase shift, which is zero at $\omega = 0$, becomes lagging with frequency to a limit of -90 degrees. Note particularly that the slope of the magnitude function is zero at $\omega = 0$, whereas that of the phase angle is finite and negative.

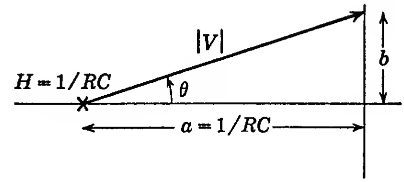


Fig. 2.9. The pole of the R-C circuit.

To clarify the mechanism of setting up and solving the equations of a circuit, an additional example will be given. Both the node and loop methods will be used. The circuit itself is the important "series-peaked" circuit shown in Fig. 2.10. Figure 2.10a is for a voltage generator, and

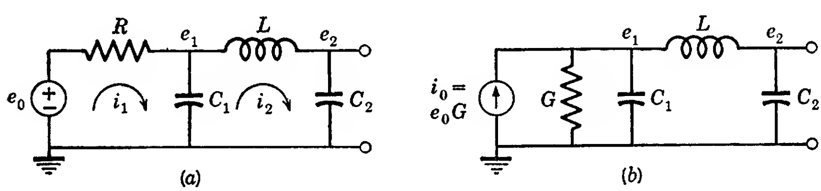


Fig. 2.10. The series-peaked circuit.

Fig. 2.10b is for a current generator, a formal exchange of sources relating the two.

The loop equations are

$$
\begin{aligned}
Ri_1 + \frac{1}{C_1}\int i_1\,dt - \frac{1}{C_1}\int i_2\,dt &= e_0 \\
-\frac{1}{C_1}\int i_1\,dt + \left(\frac{1}{C_1} + \frac{1}{C_2}\right)\int i_2\,dt + L\frac{di_2}{dt} &= 0
\end{aligned}
\tag{2.23}
$$

In terms of the complex variable p, eqs. 2.23 become

$$
\begin{aligned}
(R + 1/pC_1)i_1 - (1/pC_1)i_2 &= e_0 \\
-(1/pC_1)i_1 + (Lp + 1/pC_1 + 1/pC_2)i_2 &= 0
\end{aligned}
\tag{2.24}
$$

These equations are characteristic of a bilateral network. If driven with a vacuum tube, the system will be isolated.

The value of i_2 is given by

$$i_2 = -\frac{\Delta_{12}{}^m}{\Delta^m} e_0 = \frac{-(-1/pC_1)e_0}{\begin{vmatrix} (R + 1/pC_1) & (-1/pC_1) \\ (-1/pC_1) & (pL + 1/pC_1 + 1/pC_2) \end{vmatrix}} \tag{2.25}$$

which becomes

$$\frac{i_2}{e_0} = \frac{1/pC_1}{(R + 1/pC_1)(pL + 1/pC_1 + 1/pC_2) - (1/pC_1)^2} \tag{2.26}$$

Usually, the output voltage e_2 is to be determined rather than the current i_2. This is one reason that the node method is often simpler than the loop method. The current i_2 flows through C_2. Thus

$$e_2 = (1/C_2)\int i_2\, dt = (1/pC_2)i_2 \tag{2.27}$$

Substituting eq. 2.27 in 2.26, the transfer function is obtained as

$$\frac{e_2}{e_0} = \frac{1/(p^2C_1C_2)}{(R + 1/pC_1)(pL + 1/pC_1 + 1/pC_2) - (1/pC_1)^2} \tag{2.28}$$

Multiplying the numerator and denominator by p^2 and simplifying

$$\frac{e_2}{e_0} = \frac{1/(RLC_1C_2)}{p^3 + (1/RC_1)p^2 + [(C_1 + C_2)/(LC_1C_2)]p + 1/(RLC_1C_2)} \tag{2.29}$$

Functionally, eq. 2.29 has the form

$$\frac{e_2}{e_0} = \frac{H}{(p - p_1)(p - p_2)(p - p_3)} \tag{2.30}$$

where p_1, p_2, and p_3 all have negative real parts. The three possible p-z plots of the function are shown in Fig. 2.11. For the p-z plots of Figs.

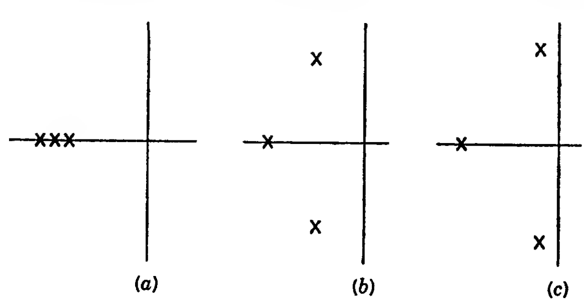


Fig. 2.11. Possible pole locations for the series-peaked circuit.

2.11*a* and *b*, the amplification of the system falls off with frequency; that is, the ratio of output voltage to input voltage is appreciable only at the lower frequencies and the function is said to be of the "low-pass" type. The gain at $\omega = 0$ (the value of e_2/e_0) is unity. For the p-z plot of Fig. 2.11*c*, the gain is appreciable only at frequencies near the complex-conjugate pole where the phasor length from that pole to the $j\omega$ axis is small. Then, the function is said to be of the "band-pass" type because it passes only a band of frequencies with relatively high gain. The amplitudes as functions of frequency for the plots of Fig. 2.11 are

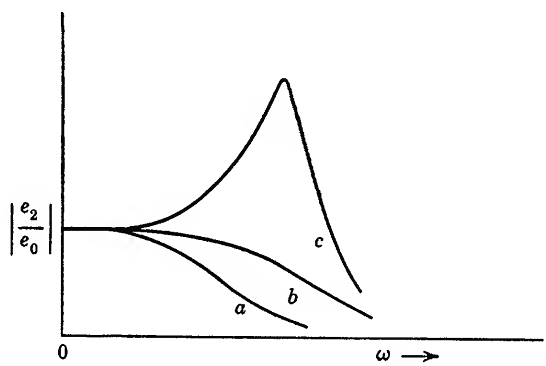


Fig. 2.12. Gain functions of the series-peaked circuit.

shown in Fig. 2.12 and may be termed "gain functions." The corresponding phase functions of frequency are shown in Fig. 2.13. For the network of this example, there is a more or less continuous transition from low-pass to band-pass behavior as the circuit element values are changed.

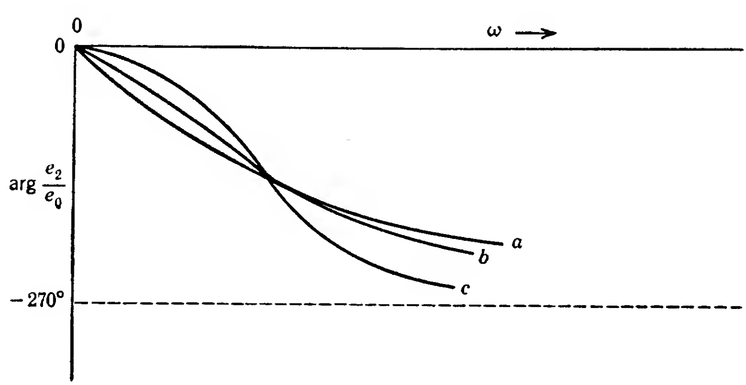


Fig. 2.13. Phase functions of the series-peaked circuit.

We shall also derive the equations for the same circuit with node equations. It will be seen that the process is somewhat simpler. With reference to the circuit of Fig. 2.10*b*, the node equations can be written as

$$
\begin{aligned}
G + C_1 \frac{de_1}{dt} + \frac{1}{L}\int e_1\,dt - \frac{1}{L}\int e_2\,dt &= i_0 = Ge_0 \\
-\frac{1}{L}\int e_1\,dt + C_2 \frac{de_2}{dt} + \frac{1}{L}\int e_2\,dt &= 0
\end{aligned}
\tag{2.31}
$$

In terms of the complex variable, these equations become

$$
\begin{aligned}
(G + pC_1 + 1/pL)e_1 - (1/pL)e_2 &= i_0 = Ge_0 \\
-(1/pL)e_1 + (pC_2 + 1/pL)e_2 &= 0
\end{aligned}
\tag{2.32}
$$

Thus

$$
e_2 = -\frac{\Delta_{12}{}^n}{\Delta^n}\,i_0 = \frac{-(-1/pL)i_0}{\begin{vmatrix} (G + pC_1 + 1/pL) & (-1/pL) \\ (-1/pL) & (pC_2 + 1/pL) \end{vmatrix}}
\tag{2.33}
$$

Simplifying and noting that the driving current i_0 is equal to a voltage e_0G, eq. 2.29 is obtained.

2.4 The partial-fraction expansion

A rational fraction is said to contain "simple" or "distinct" poles when no multiple-order poles exist. Such functions may be expanded into a number of separate terms according to eqs. 2.34 in which the C's are real constants and the A's are either complex or real constants

$$
\begin{aligned}
F(p) &= H\,\frac{p^m + a_{m-1}p^{m-1} + \cdots + a_1 p + a_0}{(p - p_1)(p - p_2)\cdots(p - p_n)} \\
&= C_0 + C_1 p + \cdots + C_q p^q + \frac{A_1}{p - p_1} + \frac{A_2}{p - p_2} + \cdots + \frac{A_n}{p - p_n}
\end{aligned}
\tag{2.34}
$$

The polynomial $C_0 + C_1 p + \cdots$ will exist only if the degree of the numerator of the rational fraction is equal to or larger than the degree of the denominator. These coefficients can be found by dividing the

denominator into the numerator. For example

$$\frac{p^2 + a_1 p + a_0}{p + b_0} = p + \frac{(a_1 - b_0)p + a_0}{p + b_0}$$
$$= p + (a_1 - b_0) + \frac{a_0 - (a_1 - b_0)b_0}{p + b_0} \tag{2.35}$$

Note that for $m = n$, $C_1 = C_2 = \cdots = 0$ and C_0 is found as the value of $F(p)$ at $p = \infty$.

The coefficient A_i in eqs. 2.34 can be found by multiplying both sides of the equation by $p - p_i$ and evaluating at $p = p_i$. Then, all other factors drop out, leaving only A_i

$$A_i = [(p - p_i)F(p)]_{p=p_i} \tag{2.36}$$

For multiple-order poles, a slightly different procedure must be employed. Let us assume that $p_1 = p_2$ in eqs. 2.34 in order to study the relations for a second-order pole. The partial-fraction expansion is

$$F(p) = C_0 + C_1 p + \cdots + C_q p^q + \frac{A_1}{(p - p_1)^2}$$
$$+ \frac{A_2}{p - p_1} + \frac{A_3}{p - p_3} + \cdots + \frac{A_n}{p - p_n} \tag{2.37}$$

The coefficient A_1 is found from

$$A_1 = [(p - p_1)^2 F(p)]_{p=p_i} \tag{2.38}$$

and the other coefficients with the exception of A_2 are evaluated as before. It can be observed that A_2 can be determined from

$$A_2 = \frac{1}{1!}\left[\frac{d}{dp}[(p - p_1)^2 F(p)]\right]_{p=p_1} \tag{2.39}$$

which can easily be verified by the reader.

The general method for treating higher order poles will now be described. Assume that $p_1 = p_2 = p_3$ in eqs. 2.34 so that a third-order pole is given. Then the expansion is

$$F(p) = C_0 + C_1 p + \cdots + C_q p^q + \frac{A_1}{(p - p_1)^3}$$
$$+ \frac{A_2}{(p - p_1)^2} + \frac{A_3}{p - p_1} + \frac{A_4}{p - p_4} + \cdots + \frac{A_n}{p - p_n} \tag{2.40}$$

The constants C_0, C_1, $\cdots$ and A_4, A_5, $\cdots$, A_n are evaluated as before. The constants A_1, A_2, and A_3 are obtained from

$$\begin{aligned} A_1 &= [(p - p_1)^3 F(p)]_{p=p_1} \\ A_2 &= \frac{1}{1!}\left[\frac{d}{dp}[(p - p_1)^3 F(p)]\right]_{p=p_1} \\ A_3 &= \frac{1}{2!}\left[\frac{d^2}{dp^2}[(p - p_1)^3 F(p)]\right]_{p=p_1} \end{aligned} \tag{2.41}$$

The method just described can be extended to poles of any order; the expressions needed for evaluating the constants associated with poles of higher order than three should be obvious from eqs. 2.41.

The coefficient of $1/(p - p_1)$ in the partial-fraction expansion has a very special definition. In reference to eqs. 2.34, the "residue" is given by

$$A_i = \text{Residue of pole } p_i \tag{2.42}$$

The concept of residues is of considerable value in transient analysis. The residue for simple poles should be clear. For multiple-order poles, it is the coefficient of $1/(p - p_i)$ to the first power.

Many of the details for obtaining a partial-fraction expansion can best be explained by means of an example. For simplicity, we shall concern ourselves with a function having simple poles as

$$\begin{aligned} F(p) &= \frac{1}{p^3 + 2p^2 + 2p + 1} = \frac{1}{(p + 1)(p^2 + p + 1)} \\ &= \frac{A}{p + 1} + \frac{B}{p + [1 + j(3)^{1/2}]/2} + \frac{C}{p + [1 - j(3)^{1/2}]/2} \end{aligned} \tag{2.43}$$

To evaluate A, we multiply both sides of eqs. 2.43 by $p + 1$ and evaluate at $p = -1$ to get

$$A = \left(\frac{1}{p^2 + p + 1}\right)_{p=-1} = 1 \tag{2.44}$$

Similarly

$$B = \left[\frac{1}{(p + 1)\{p + [1 - j(3)^{1/2}]/2\}}\right]_{p=[-1-j(3)^{1/2}]/2} = \frac{-2}{3 + j(3)^{1/2}} \tag{2.45}$$

The quantities B and C are related to a pair of complex-conjugate poles. When partial fractions relating to such poles are combined, the result must not contain numbers with imaginary parts or else the pair of terms will not be conjugate analytic. Therefore, it is necessary that

$$C = B^* = \frac{-2}{3 - j(3)^{1/2}} \tag{2.46}$$

which is a general requirement that can be used to reduce the work of finding the partial-fraction expansion.

Sometimes, pairs of terms related to pairs of complex-conjugate poles are combined. This gives for the example

$$\frac{1}{p^3 + 2p^2 + 2p + 1} = \frac{1}{p+1} - \frac{p}{p^2 + p + 1} \tag{2.47}$$

2.5 The continued-fraction expansion

Another important expansion having a characteristic form is the continued-fraction expansion. It is obtained by successive division and inversion of a rational fraction. Consider, for example, the function given by eq. 2.48

$$F(p) = \frac{p^4 + p^3 + 3p^2 + 2p + 1}{p^3 + p^2 + 2p + 1} \tag{2.48}$$

Dividing the denominator into the numerator beginning with the highest powers in p and inverting the remainder, we get

$$F(p) = p + \cfrac{1}{\cfrac{p^3 + p^2 + 2p + 1}{p^2 + p + 1}} \tag{2.49}$$

The division and inversion of the remainder can be repeated for the rational fraction appearing in the denominator of the second term of eq. 2.49 to obtain

$$F(p) = p + \cfrac{1}{p + \cfrac{1}{\cfrac{p^2 + p + 1}{p + 1}}} \tag{2.50}$$

Repeating the process again

$$F(p) = p + \cfrac{1}{p + \cfrac{1}{p + \cfrac{1}{p + 1}}} \tag{2.51}$$

A rational fraction may always be expressed in terms of a continued-fraction expansion. However, the expansion may not always terminate and/or the separate terms may have fractional powers of p or $1/p$, in which case the expansion is of little practical value.

A more generalized kind of continued-fraction expansion is one having the form

$$F(p) = F_1(p) + \cfrac{1}{F_2(p) + \cfrac{1}{F_3(p) + \cfrac{1}{F_4(p) + \cdots}}} \tag{2.52}$$

where $F_1, F_2, F_3, \cdots$ are functions of p. If these functions are in themselves rational fractions and subject to the laws of input immittances to be set forth in Chapter 3, the expansion is a useful one. (It is related to the general ladder network.) The expansion of eq. 2.51 is a special case of the more general expansion of eq. 2.52.

When confronted with a rational fraction that is to be expressed as a continued fraction, trial and error can be applied. The expansion can often be found, particularly if it corresponds to a broad class of networks referred to as ladders. Trial and error in the inversion and division process is often required. For example, the division of the function $(p^2 + p + 1)/(p^2 + 1)$ can be started in any of the following manners:

$$\begin{array}{ll} p^2 + 1 \,\overline{\big)\, p^2 + p + 1} & \quad 1 + p^2 \,\overline{\big)\, 1 + p + p^2} \\ \\ p^2 + p + 1 \,\overline{\big)\, p^2 + 1} & \quad 1 + p + p^2 \,\overline{\big)\, 1 + p^2} \end{array} \tag{2.53}$$

Alternately, two or more terms can sometimes be divided at the same time as

$$\frac{p^2 + p + 1}{p^2 + 1} = \frac{(p^2 + 1) + p}{p^2 + 1} = 1 + \frac{p}{p^2 + 1} \tag{2.54}$$

One trial may yield an expansion that has no physical interpretation, whereas another trial might have meaning. It is also possible that no meaningful expansion can be found.

The continued-fraction expansion also has a valuable application in determining whether or not a polynomial is a Hurwitz polynomial. A Hurwitz polynomial was defined as one having zeros only to the left of the $j\omega$ axis. Along the positive $j\omega$ axis, the phase angle of such a polynomial must be a monotonically increasing function of frequency, which is a property that should be obvious from the p-plane plot of the zeros. Thus, the phase angle starts at zero and progresses through 90, 180, 270 degrees, and so on, depending upon the degree of the polynomial. Ultimately, the phase angle must reach n times 90 degrees, where n is the total number of zeros (the degree of the polynomial). The tangent of the phase angle must therefore be zero at zero frequency, next be infinite, next zero, and so on.

Consider a Hurwitz polynomial $F(p)$ at $p = j\omega$, that is, along the positive $j\omega$ axis

$$\begin{aligned} F(p) &= a_0 + a_1 p + a_2 p^2 + \cdots + a_n p^n \\ F(j\omega) &= F_1(\omega) + jF_2(\omega) \end{aligned} \tag{2.55}$$

where

$$\begin{aligned} F_1(\omega) &= [\mathrm{Ev}\ F(p)]_{p=j\omega} \\ jF_2(\omega) &= [\mathrm{Od}\ F(p)]_{p=j\omega} \end{aligned} \tag{2.56}$$

The phase angle θ of $F(j\omega)$ is

$$\theta = \arg F(j\omega) = \tan^{-1} \frac{F_2(\omega)}{F_1(\omega)} \tag{2.57}$$

Thus

$$j \tan \theta = \left(\frac{a_1 p + a_3 p^3 + a_5 p^5 + \cdots}{a_0 + a_2 p^2 + a_4 p^4 + \cdots} \right)_{p=j\omega} \tag{2.58}$$

It is convenient to define the "tangent function" of p from eq. 2.58 as

$$F_p(\tan \theta) = \frac{\mathrm{Od}\ F(p)}{\mathrm{Ev}\ F(p)} = \frac{a_1 p + a_3 p^3 + \cdots}{a_0 + a_2 p^2 + \cdots} \tag{2.59}$$

Since the phase angle of a Hurwitz polynomial is an ever-increasing function of ω, all the p-z of the tangent function must lie on the $j\omega$ axis, must be simple, and must alternate such that a zero occurs between two poles and conversely. Further, if n is odd, there will be n zeros and $n - 1$ poles, whereas if n is even, there will be n poles and $n - 1$ zeros. Examples on the p plane for $n = 5$ and $n = 6$ are shown in Fig. 2.14.

In Chapter 3, a function such as that of Fig. 2.14 will be shown to be a reactance function and *always* obtainable as the input immittance of a reactive network (only inductance and capacitance). This means that the continued-fraction expansion of the tangent function of a Hurwitz

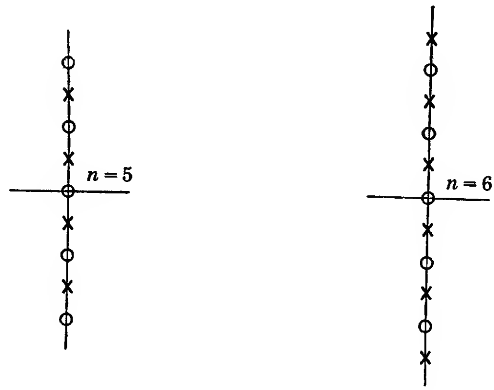


Fig. 2.14. Tangent functions for $n = 5$ and $n = 6$.

polynomial can always have one of the forms of eqs. 2.60, perhaps in addition to others

$$
\begin{gathered}
b_1p + \cfrac{1}{b_2p + \cfrac{1}{b_3p + \cfrac{1}{b_4p + \cdots}}} \\
c_1/p + \cfrac{1}{c_2/p + \cfrac{1}{c_3/p + \cfrac{1}{c_4/p + \cdots}}}
\end{gathered}
\tag{2.60}
$$

All the b_i and c_i in eqs. 2.60 must be positive, real, and greater than zero if the original polynomial has all its zeros within the left half-plane; otherwise the polynomial will not be a Hurwitz polynomial (because it cannot be represented with a physical reactance network). If any of the b_i or c_i are negative, then one or more zeros will lie in the right half-plane. If one of the b_i or c_i is zero, then one pair of complex-conjugate zeros will lie exactly on the $j\omega$ axis.

Any polynomial in p can be tested as described in order to see whether or not it is "Hurwitz." It may be necessary to try division starting from both the lowest and highest powers of p in order to get the answer. As

an example, consider the following illustration

$$
\begin{aligned}
F(p) &= p^3 + 2p^2 + 2p + 1 \\
\frac{\mathrm{Od}\, F(p)}{\mathrm{Ev}\, F(p)} &= \frac{p^3 + 2p}{2p^2 + 1} = \frac{p}{2} + \cfrac{1}{4p/3 + \cfrac{1}{3p/2}}
\end{aligned}
\tag{2.61}
$$

From eqs. 2.61, it can be seen that the polynomial is a Hurwitz polynomial. However, had we divided the same function starting from the lower powers of p, we would have obtained eq. 2.62, which is not very promising

$$
\frac{2p + p^3}{1 + 2p^2} = 2p + \cfrac{1}{\cfrac{2p^2 + 1}{-3p^3}} \tag{2.62}
$$

Problems

1. Prove that the amplitude and phase of a conjugate-analytic function are even and odd functions of frequency, respectively, along the $j\omega$ axis of the p plane.

2. Prove that a conjugate-analytic polynomial has purely real coefficients.

3. A function has zeros at $-1 \pm j1$ and -3. It has poles at -2 and $-1 \pm j5$.

a. Plot the p-z on the p plane.

b. Sketch the amplitude and phase of the function along the positive $j\omega$ axis of the p plane.

c. Form the factored expression of the function.

d. Put the function in the form of two polynomials in p.

e. Obtain the new function giving the square of the magnitude of the original function along the $j\omega$ axis.

4. Given the polynomial $p^3 + 2p^2 + 2p + 1$: Form the function that is the ratio of the odd and even parts of the polynomial and plot the locations of its p-z.

5. The function $\sin p$ has zeros at multiples of $p = \pm\pi$. Sketch the positions of these zeros. Assuming that in addition to a constant multiplier the locations of the zeros are sufficient to describe the function uniquely, prove that the infinite-product expansion for $\sin p$ is

$$
\begin{aligned}
\sin p &= p \prod_{k=1}^{\infty} \left(1 - \frac{p^2}{k^2\pi^2}\right) \\
&= p(1 - p^2/\pi^2)(1 - p^2/4\pi^2) \cdots
\end{aligned}
$$

6. Repeat Prob. 5 for the following functions:

$$
\cos p = \prod_{k=1}^{\infty} \left(1 - \frac{p^2}{(2k-1)^2(\pi/2)^2}\right)
$$

$$
\sinh p = p \prod_{k=1}^{\infty} \left(1 + \frac{p^2}{k^2\pi^2}\right)
$$

$$
\cosh p = \prod_{k=1}^{\infty} \left(1 + \frac{p^2}{(2k-1)^2(\pi/2)^2}\right)
$$

7. Plot the p-z of $\tan p$ and $\tanh p$ and give their infinite-product expansions.

8. The exponential $e^{ja\omega}$ has a phase angle $a\omega$; that is, $\arg(e^{ja\omega}) = a\omega$. Show the factored infinite-product expansion for $\tan a\omega$ and plot the locations of the p-z of this function.

9. Given a function having n poles and m zeros, all on or left of the $j\omega$ axis, with q of the m zeros at the origin: What will be the phase shift for $p = j\omega$ very small and very large? What will be the shape of the magnitude of the function for $p = j\omega$ very small and very large? The answers asked for include all values of q from zero to m.

10. A polynomial is $p^3 + 2p^2 + 2p + 1$. For $p = j\omega$ and using ordinary complex algebra, obtain an expression for the tangent of the phase angle of the polynomial. Form the tangent function of p of this polynomial and plot its p-z. From this plot, write down the tangent function evaluated for $p = j\omega$. What is the minor difference between the expressions for the tangent as obtained in the two ways described?

11. Given the function $F(p) = (p + 1)/(p^2 + p + 1)$: Plot the magnitude and phase of this function from the p-plane plot of its p-z along the entire $j\omega$ axis using a protractor and a pair of dividers.

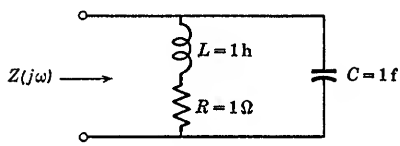


Fig. P.12.

12. Plot the magnitude and phase angle of the impedance of the circuit of Fig. P.12 for all values of ω using conventional complex algebra. Compare the results to those of Prob. 11.

13. Plot the p-z of the function giving the square of the magnitude of the function of Prob. 11 along the $j\omega$ axis. Directly from the positions of these p-z, obtain this new function as the ratio of two polynomials in p.

14. Find the partial-fraction expansions of the following

$$\frac{p^2 + 1}{p(p^2 + 3p + 1)} \qquad \frac{p^2 + 1}{p(p^2 + 2p + 1)}$$

$$\frac{p^2 + 1}{p(p^2 + p + 1)} \qquad \frac{p(p^2 + 1)}{p^2 + 3p + 1}$$

$$\frac{1}{(p + 1)(p + 2)(p + 3)} \qquad \frac{p}{(p + 1)(p^2 + p + 1)^2}$$

15. Obtain the continued-fraction expansion of the following function starting from the highest powers of p such that each term will be positive and proportional to p, $1/p$, or a constant

$$\frac{p^5 + p^4 + 4p^3 + 3p^2 + 3p + 1}{p^4 + p^3 + 3p^2 + 2p + 1}$$

16. Obtain the continued-fraction expansion of the following function with terms that are positive and proportional to p, $1/p$, or a constant. Trial and error may have

to be used

$$\frac{p^2LGC + pC + G}{p^2LC + 1}$$

17. Test the following polynomials to see if they are Hurwitz

$$p^3 + 5p^2 + 8p + 4$$
$$p^3 + p^2/2 + p + 1$$
$$p^4 + 2p^3 + 2p^2 + p + 1$$
$$p^3 + 2 \times 10^6p^2 + 10^{12}p + 10^{18}$$
$$0.5p^3 + 0.5 \times 10^6p^2 + 0.5 \times 10^{12}p + 10^{18}$$

18. The ratio of two reactance polynomials is important to the theory of input immittances of L-C networks. The partial-fraction expansions of such functions can be simplified to a certain extent. Show that when the number of p-z differ by precisely unity, when all p-z lie on the $j\omega$ axis, when all p-z are simple, and when all p-z (except for one at the origin) occur in complex-conjugate pairs, the expansion will *always* have one of the two forms

$$I = \begin{cases} H\dfrac{(p^2 + a_1^2)(p^2 + a_2^2)\cdots}{p(p^2 + b_1^2)(p^2 + b_2^2)\cdots} = \dfrac{A_0}{p} + A_\infty p + \dfrac{A_1 p}{p^2 + b_1^2} + \cdots \\ H\dfrac{p(p^2 + a_1^2)(p^2 + a_2^2)\cdots}{(p^2 + b_1^2)(p^2 + b_2^2)\cdots} = A_\infty p + \dfrac{A_1 p}{p^2 + b_1^2} + \cdots \end{cases}$$

19. Show that the constants A_0 and A_∞ of Prob. 18 are

$$A_0 = (pI)_{p=0} = H\frac{a_1^2a_2^2\cdots}{b_1^2b_2^2\cdots} \qquad A_\infty = \left(\frac{I}{p}\right)_{p=\infty} = \begin{cases} H \\ \text{or} \\ 0 \end{cases}$$

20. Show that the constants A_k of Prob. 18 for $k > 0$ are

$$A_k = \left(\frac{(p^2 + b_k^2)I}{p}\right)_{p=\pm jb_k}$$

21. If a conjugate-analytic function is expanded into partial fractions, then for each pair of complex-conjugate poles, there will occur two terms as

$$\frac{A_1}{p + \alpha + j\beta} + \frac{A_2}{p + \alpha - j\beta}$$

Show that A_1 and A_2 must be complex conjugates.

22. Prove the following alternate method for evaluating the constant multiplier associated with one of the terms of a partial-fraction expansion: If $B(p)$ is a polynomial having a simple pole at $p = p_k$ which is not also a root of the polynomial $A(p)$, then

$$\left((p - p_k)\frac{A(p)}{B(p)}\right)_{p=p_k} = \left(\frac{A(p)}{\dfrac{dB(p)}{dp}}\right)_{p=p_k}$$

3

The Poles and Zeros of Networks

It has been observed that all input and transfer immittances and transfer functions can be interpreted as rational functions of the complex variable and as such may be expressed as the ratio of two polynomials in p having real coefficients. Thus, p-z will be either purely real or occur in complex-conjugate pairs. It was also concluded that, along the imaginary axis, the magnitude and phase of immittances are even and odd functions of frequency respectively.

In addition to these restrictions regarding network functions, several other rather stringent laws apply to the permissible locations of the p-z. It is the purpose of this chapter to set forth these laws in terms of what can and cannot be done with electric circuits.

The p-z of networks often have rather characteristic positions on the p plane. This leads to the formulation of definitions of network types based on their locations.

The close relationship between p-z locations and transient response naturally leads to a discussion of transient analysis. The development here is based on the Heaviside method. General driving functions and some initial conditions will be treated, and the intimate relationship between Heaviside and Laplace methods will be pointed out.

3.1 The p-z of transfer functions

Let a transfer function be expressed as a partial-fraction expansion and let it be assumed that all the poles are simple. Then

$$\frac{\text{Output}}{\text{Input}} = \frac{e_n}{e_0} = C_0 + C_1 p + C_2 p^2 + \cdots + \frac{A_1}{p - p_1} + \frac{A_2}{p - p_2} + \cdots + \frac{A_n}{p - p_n} \tag{3.1}$$

Since the system is assumed to be linear, the principle of superposition can be utilized to show that the output is made up of a linear combina-

tion of several inputs as

$$e_n = C_0 e_0 + C_1 p e_0 + \cdots + \frac{A_1}{p - p_1} e_0 + \cdots + \frac{A_n}{p - p_n} e_0 \tag{3.2}$$

The complex variable p can be interpreted to be the differential operator d/dt. Then, eq. 3.2 can be interpreted as a number of simple differential equations. Let us examine a typical term giving a partial output δe_n, which is the output that results when only *one* of the constants A_j or C_k is other than zero

$$\delta e_n = \frac{A_k}{p - p_k} e_0 \tag{3.3}$$

In terms of differentials, this equation becomes

$$\frac{d\delta e_n}{dt} - p_k \delta e_n = A_k e_0 \tag{3.4}$$

This simple kind of differential equation has a solution made up of two parts, one for which e_0 is zero and the other for e_0 not zero. We shall be most interested in the solution for vanishing e_0. Then

$$\frac{d\delta e_n}{dt} - p_k \delta e_n = 0 \tag{3.5}$$

Rearranging

$$\frac{d\delta e_n}{\delta e_n} = p_k \, dt \tag{3.6}$$

Integrating both sides of eq. 3.6 and neglecting the constants of integration

$$\ln \delta e_n = p_k t \tag{3.7}$$

where ln signifies the natural logarithm. In terms of an exponential, eq. 3.7 is

$$\delta e_n = e^{p_k t} = e^{(\sigma_k t + j\omega t)} \tag{3.8}$$

If σ_k is positive, δe_n grows exponentially with time even though the input to the network is negligible. This is contrary to the law of conservation of energy and we conclude: *The poles of stable transfer functions can only occur on or to the left of the $j\omega$ axis.* In other words, the right half-plane is a prohibited region for the poles of stable transfer functions.

Let us now interpret p as the imaginary variable $j\omega$. Consider one of the partial outputs of eq. 3.2 given by the coefficient C_k

$$\delta e_n = C_k p^k e_0 \rightarrow C_k (j\omega)^k e_0 \tag{3.9}$$

As the frequency of e_0 tends to infinity, the amplitude of this partial output tends to infinity. Clearly, this is also contrary to the principle of conservation of energy; hence it must be concluded that $C_1 = C_2 = C_3 = \cdots = 0$. Thus: *The number of zeros of a stable transfer function cannot be greater than the number of poles.*

If the source driving a network is ideal in the sense that its internal impedance is precisely zero or infinite, depending upon whether it is a voltage source or a current source respectively, then the source can be

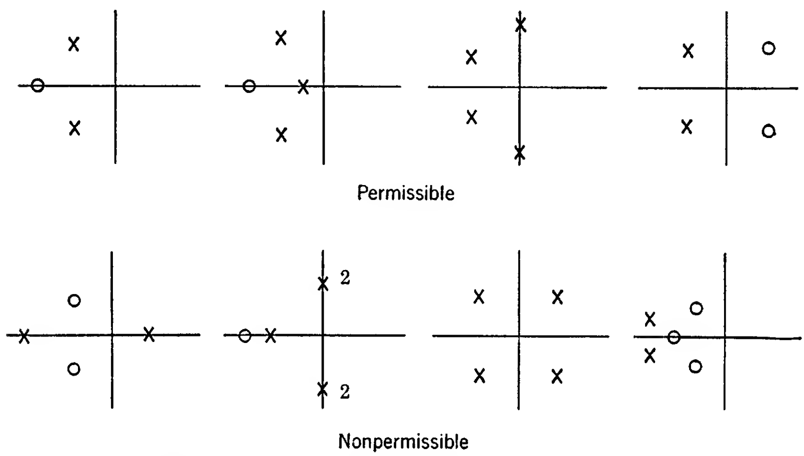


Fig. 3.1. P-z locations of stable transfer functions.

made to deliver infinite power. In this event, the transfer function can have one (but only one) more zeros than poles. However, ideal sources do not occur in nature but only exist as an approximation. As long as the nature of the approximation is clear to the network designer, it is permissible to employ ideal sources.

A somewhat analogous treatment of the differential equations previously considered yields a relevant partial output that increases with time if multiple-order poles lie on the imaginary axis. Thus: *Poles of a stable transfer function lying on the imaginary axis must be simple.*

The rules that have been set down so far can be illustrated by the p-plane plots of Fig. 3.1.

It is important to note that no restrictions regarding the positioning of the zeros have been given except that they, like the poles, must occur in complex-conjugate pairs (if complex), and they cannot be too numerous.

It is of interest to consider the character of a system having poles in the right half-plane. Such systems are unstable since an output growing with time will be present, even though the input is zero. Consider

a complex pole lying at $\sigma_k + j\omega_k$ where σ_k is positive. From eq. 3.8, we see that an oscillation occurs at the frequency ω_k whose magnitude grows exponentially with time. Of course, the magnitude must eventually be limited by the nonlinearities which exist in any practical system. It can be concluded that a transfer function showing a pair of complex-conjugate poles in the right half-plane can be made to produce sinusoidal oscillations at a frequency given approximately by the imaginary part of the pole position. The rapidity with which the oscillation builds up after the system is initially energized is proportional to the real part of the pole position.

When a pole lies on the positive real axis of the p plane, the output will rise exponentially with time without a superimposed sinusoidal variation. Then, relaxation oscillations as in a multivibrator may result.

It should be clear that, if a pair of complex-conjugate poles lies exactly on the $j\omega$ axis, an oscillation in the network, once initiated, will continue undamped for all time at the frequency of the pole position, even though the input is removed. Poles exactly on the $j\omega$ axis represent the critical condition at which oscillation commences.

Sometimes a transfer function may have two or more pairs of complex-conjugate poles in the right half-plane, indicating an ability to support oscillations at different frequencies. Oscillation will normally occur as governed by the pair of poles farthest right of the imaginary axis relative to the imaginary distance from the origin.

3.2 The p-z of bilateral input immittances

Bilateral networks have no sources of energy, and hence must be stable. Consequently, the transfer impedance or admittance can have poles in the left half-plane only. Simple poles may also exist on the imaginary axis. The number of zeros cannot exceed the number of poles by more than one.

Let us now consider the rules regarding the input immittance of a bilateral network. Assume that we have a bilateral network accessible from a pair of terminals. First, a voltage generator e_0 having a source impedance Z_s is caused to drive the network as in Fig. 3.2*a*. Then, the *same* network is driven with a current generator i_0 having a source admittance Y_s as shown in Fig. 3.2*b*. The input current and voltage resulting from the application of the driving voltage and current, respectively, are

$$i = \frac{e_0}{Z_s + Z_{\text{in}}} \qquad e = \frac{i_0}{Y_s + Y_{\text{in}}} \tag{3.10}$$

If we let $Z_s \to 0$ for the voltage generator and $Y_s \to 0$ for the current generator, we get $i = Y_{\text{in}}e_0$ and $e = Z_{\text{in}}i_0$ respectively. If arbitrary

driving functions are not to produce effects in the network that increase indefinitely with time after the driving function is removed, there can be no poles of Y_{in} in the right half-plane, and poles, if any, on the $j\omega$ axis must be simple. Similarly, the poles of Z_{in} cannot occur in the right half-plane. Since the poles of Z_{in} are in the zeros of Y_{in} and conversely, we conclude: *Neither the poles nor the zeros of the input immittance*

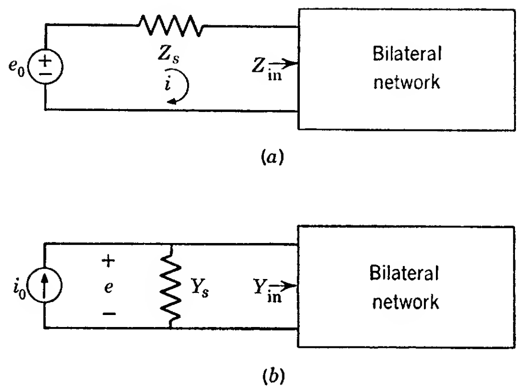


Fig. 3.2. Bilateral network with source immittance.

to a bilateral network can lie in the right half-plane, and p-z on the imaginary axis must be simple.

Consider the parallel-resonant circuit (R, L, and C all in parallel) driven with a current i. The voltage to current ratio is

$$\frac{e}{i} = \frac{p/C}{p^2 + p/RC + 1/LC} \tag{3.11}$$

The poles of this circuit are normally in the left half-plane. However, if R is negative, they will lie in the right half-plane. This simple example demonstrates that the input immittance of a bilateral network cannot have a negative real part (otherwise, self-oscillations could be produced).

We have already established that the p-z of a bilateral input immittance must lie in the left half-plane and that p-z along the $j\omega$ axis must be simple. In addition, this tells us that the real part of the immittance evaluated along the $j\omega$ axis cannot be negative. This means that the phase shift can never exceed 90 degrees lagging or leading which, from p-plane sketches studied at very high frequencies, is seen to mean: *The number of p-z of the input immittance of a bilateral network can differ at most by unity.* The detailed positioning of these p-z must be such that

the phase shift along the $j\omega$ axis of the p-plane is forever bounded by ± 90 degrees.

In the preceding, we have deduced the so-called "positive-real" nature of bilateral input immittances. It turns out that positive realness is not only a necessary but is also a sufficient condition for a function to represent a physical input immittance.

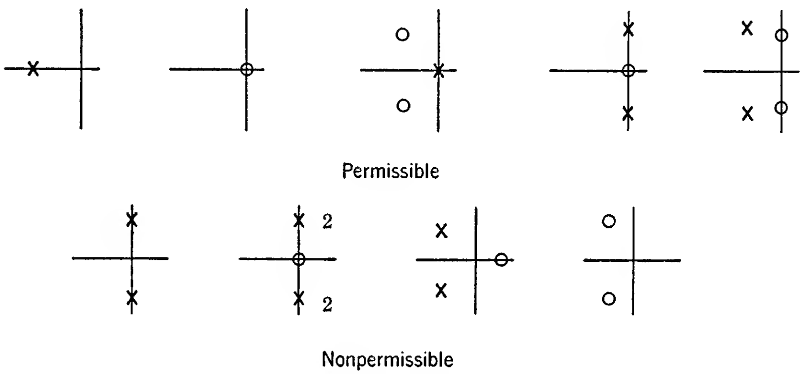


Fig. 3.3. P-z of bilateral input immittances.

As an aid to the visualization of permissible and nonpermissible p-z locations for input immittances, the reader may refer to Fig. 3.3. The restrictions imposed on these p-z are notable.

3.3 The p-z of reactive networks

Let it be assumed that a bilateral network contains only L, C, and M; that is, resistance is entirely absent. Then absolutely no power can be dissipated in the network and sinusoidal voltages and currents relating to any element in the network must always be separated by a 90-degree phase angle.

In particular, consider the input immittance to such a network. The phase angle between the applied sinusoidal voltage and resulting current must always be exactly ± 90 degrees; if this were not so, power would be dissipated. By examining the p-z on the p plane, it can be observed that such phase angles are given for all ω only if *all the p-z lie exactly on the imaginary axis.* As determined earlier, such p-z *must be simple.*

In order for the phase shift to be 90 degrees at low frequencies, a simple pole or zero must occur at the origin. The presence of this pole or zero means that *the total number of p-z must always differ by unity.*

The p-z of the input immittance of an L-C network occur on the $j\omega$ axis and are simple. This means that an oscillation initiated at the fre-

quency of a pole will continue undamped forever even though the driving function initially producing the oscillation is removed. An oscillation continuing forever in an undriven bilateral circuit is not an entirely satisfying picture. We know that perfectly lossless inductors and capacitors cannot be built. Further, there will always be some radiation, no matter how small, from any circuit containing an oscillatory voltage. It must be concluded that an oscillatory disturbance in an *L-C* circuit must eventually decay to zero. This means that the poles can never be located exactly on the $j\omega$ axis but must actually occur slightly to the left of the $j\omega$ axis. Now consider the consequence if two such poles

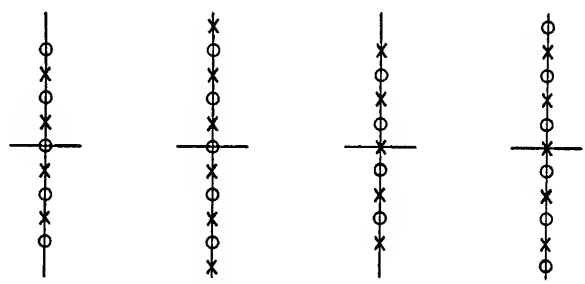


Fig. 3.4. Input immittances of reactive networks.

near the $j\omega$ axis occurred next to each other without an intervening zero. Then, as the frequency variable was moved along the $j\omega$ axis past the poles, two successive 180-degree phase shifts having the same sign would occur and the phase angle of the input immittance could not but exceed ± 90 degrees, implying the presence of a negative resistance or conductance component. It can be concluded that poles of a bilateral *L-C* input immittance must be separated by a zero. It can be seen from a similar argument that the existence of two adjacent zeros without an intervening pole is likewise impossible. The conclusion must be that *the p-z of a bilateral L-C input immittance must alternate such that a pole occurs between two zeros and conversely.*

Basically, the p-z plot of an input immittance of a reactance function can have only one of the four forms shown in Fig. 3.4. These plots are the same as those applicable to the tangent function of a Hurwitz polynomial (or its reciprocal); hence, *the tangent function of a Hurwitz polynomial is always realizable as the input immittance of a reactive network.*

3.4 The p-z of *R-C* and *R-L* networks

In *R-L* and *R-C* networks, there is no mechanism for energy interchange between different kinds of reactive elements. Consequently, we

would not expect oscillatory partial outputs to be indicated in the partial-fraction expansion of the transfer immittance. It can therefore be suspected that the p-z of the input immittance lie on the negative real axis of the p plane. In order that the phase angle not exceed 90 degrees, these p-z must not only be simple but must alternate such that a pole occurs between two zeros and conversely. The number of poles or zeros at the origin cannot exceed unity because the origin is part of the $j\omega$ axis. Of course, the total number of p-z can differ at most by unity.

At zero frequency where capacitors are all open circuits, an R-C circuit will look like either an open circuit or a finite resistance; hence, the impedance can have either a pole or nothing at the origin. The nearest thing to the origin must always be a pole. At infinite frequencies where

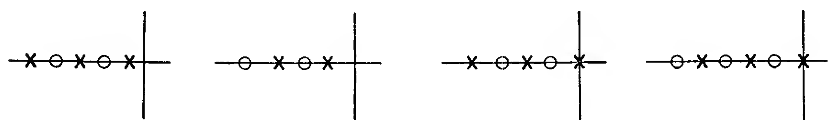


Fig. 3.5. Input impedances of R-C networks. Input admittances of R-L networks.

capacitors are all short circuits, an R-C network can have an input impedance that is either zero or finite but never infinite. These various properties can be used to find the possible p-plane sketches of the input impedance of an R-C network as shown in Fig. 3.5. The permissible forms for the input admittance of R-C networks can be found from Fig. 3.5 by interchanging p-z. The rules for R-C input impedances are equally applicable to input admittances of R-L networks.

Because the poles of all input and transfer immittances are the same, we conclude that transfer immittances of bilateral R-C or R-L networks have simple poles that all lie on the negative real axis of the p plane. The zero positions are not restricted. Of course, an isolated system (that is, systems with vacuum tubes operating as coupling devices) can have poles of any order, and an R-C network with feedback can have complex poles.

A somewhat intuitive justification has been given for the allowable positions of the p-z of R-C and R-L networks. Formal proofs of these properties may be found in the references.

3.5 Dissipation, capacitance, and approximations

If a network is driven with an ideal voltage or current source, the number of zeros of the transfer function can never exceed the number of poles by more than one. With practical sources, the number of zeros can never exceed the number of poles because a practical source cannot deliver unlimited amounts of power. (Note, however, that an input

immittance may be defined with an ideal source.) Practical systems usually have fewer zeros than poles; this arises from unavoidable stray capacitance between the wires and elements carrying signals which short-circuit high-frequency signals to ground. It is only when these distributed capacitances can justifiably be ignored (with a practical source) that the number of zeros can be made equal to the number of poles.

When all distributed capacitances are included, some of the p-z will be relatively far from the origin of the p plane. If their distances are large compared to the expanse of frequencies along the $j\omega$ axis that are of interest, and if they do not contribute an appreciable change of phase over this expanse of frequencies, distant p-z may be treated as if they did not exist (except for a constant multiplier). This is equivalent to ignoring the circuit elements involved. Of course, this approximation may cause a function to appear to have any number of p-z. In the whole of the p plane, there must, of course, be no more zeros than poles.

Occasionally, a function may have a pole and a zero in close proximity. If they are close compared to the frequency range of interest along the $j\omega$ axis, they can be assumed to exist at the same point. Then, the effect of the pole exactly cancels that of the zero and both may be ignored. Similarly, if several poles (or zeros) lie in relatively close proximity, certain simplifications are possible by assuming them to lie at the same point giving the equivalent of a multiple-order pole (or zero).

It was shown that the p-z of the input immittance of a reactive network lie exactly on the $j\omega$ axis. This is a result of approximation. No inductor can be built without including, even though unintentionally, an appreciable amount of resistance. Capacitors also have losses, although some high-grade units are perfect for all practical purposes. The result is that the p-z of reactive networks cannot lie on the $j\omega$ axis but must lie slightly to the left, that is, in the left half-plane.

Often, inductors and capacitors are assumed to be ideal in order to simplify the network equations to the point where the mathematics become tractable. Because of the "incidental dissipation" arising as a result of the residual resistances, the actual positions of the p-z on the p plane will be slightly left of their idealized positions. In arriving at some design, a synthesis of an immittance that neglects incidental dissipation should start from p-z positions somewhat to the right of the desired positions (if possible); dissipation automatically moves them to their desired positions. This is called "predistortion."

3.6 Resonant circuits: an example of approximation

Consider a parallel-resonant circuit with unavoidable resistance in series with the inductor. The resistance associated with the capacitor

will be neglected. The circuit and p-plane plot of the input impedance are shown in Fig. 3.6. The expression for the input impedance is

$$Z = \frac{1}{C}\left(\frac{p + R/L}{p^2 + pR/L + 1/LC}\right) \tag{3.12}$$

where it will be assumed that the poles have imaginary parts. If interest is confined to frequencies in the vicinity of the poles, and if the poles lie

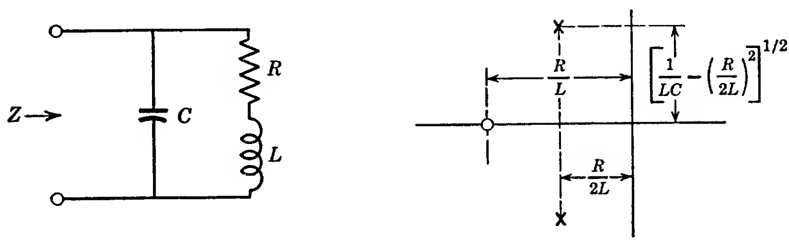


Fig. 3.6. Resonant circuit with R in L branch.

close to the $j\omega$ axis compared to their distances from the origin of the p plane, R/L in the numerator of eq. 3.12 can be neglected as an approximation. This has the effect of moving the zero to the origin. With this approximation, the circuit of Fig. 3.6 becomes equivalent to the parallel-

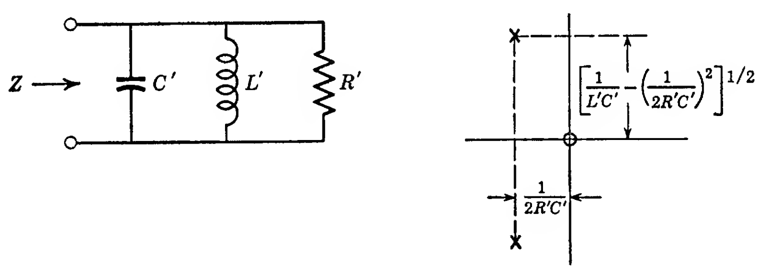


Fig. 3.7. The parallel-resonant circuit.

resonant circuit of Fig. 3.7, which has an input impedance given by

$$Z' = \frac{1}{C'}\left(\frac{p}{p^2 + p/R'C' + 1/L'C'}\right) \tag{3.13}$$

In order that the two circuits have the same pole positions (assuming $C = C'$ and $L = L'$), it is necessary that

$$R/L = 1/R'C' \quad \text{or} \quad R' = L/RC' \tag{3.14}$$

The radial distance from the origin to the pole $\omega_0 = 1/(LC)^{1/2}$ is called the resonant frequency of the circuit. At this frequency, the impedance of the parallel-resonant circuit is a maximum and the phase angle of the impedance is zero. These do not occur at precisely the same frequency when resistance occurs in series with the inductor. When R/L in the numerator of eq. 3.12 is negligible, the impedance expression for both circuits under consideration is the same, namely

$$Z = \frac{1}{C}\left(\frac{p}{p^2 + p\omega_0/Q + \omega_0^2}\right) \tag{3.15}$$

where $Q = \omega_0 L/R = \omega_0 R'C'$ is the "quality factor." In either case, ω_0/Q is twice the real part of the pole position. At $\omega = \omega_0$ for $p = j\omega$, the value of Z is $Q/C\omega_0$. At the two frequencies $\omega_0 \pm \omega_0/2Q$, the magnitude of Z is $1/(2)^{1/2}$ times that at $\omega = \omega_0$. The frequency range within these limits is $\omega_0/Q = B$, which is called the "half-power bandwidth." The number Q thus gives the "bandwidth ratio," which is the ratio of the center frequency to the bandwidth and measures the relative "selectivity" of the circuit.

The series-resonant circuit is similar to the parallel-resonant circuit. Its admittance rather than its impedance is given by eq. 3.15 except that the multiplier is $1/L$ rather than $1/C$ and $Q = \omega_0 L/R$. Thus, whatever is said about the gain and phase functions of frequency of the input impedance of the parallel-resonant circuit applies to the input admittance of the series-resonant circuit, for they are reciprocal networks.

3.7 Basic stable network transfer functions

Network transfer functions as general types fall into three major classifications: minimum-phase, nonminimum-phase, and all-pass functions. These classes of functions can readily be defined by observing their p-z locations.

Consider the four p-z plots of Fig. 3.8, where the poles in all cases are the same. All four functions have identical amplitude-versus-frequency characteristics. However, the amount of phase shift with frequency,

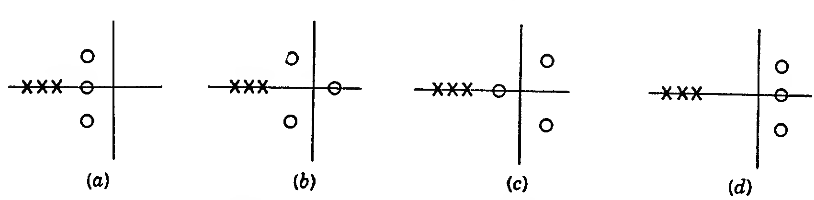


Fig. 3.8. Four possible zero locations.

normalized so that all four systems have zero phase shift at zero frequency, is a minimum for Fig. 3.8*a*, as can be seen by observing the behavior of the phasors from the zeros as the frequency increases. Thus, we can define a minimum-phase transfer function as one which has zeros in the left half-plane only. Figure 3.8*a* is minimum phase; Figs. 3.8*b*, *c*, and *d* are not. For a function to be nonminimum phase, the

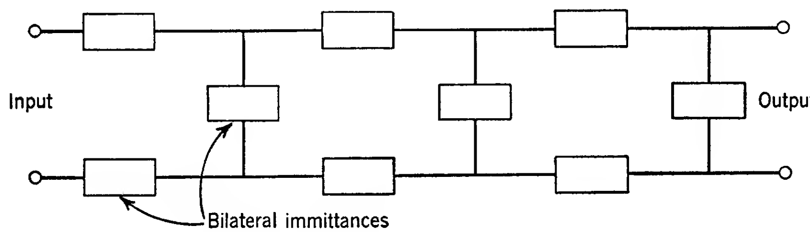


Fig. 3.9. The ladder network.

physical system must possess means for the input signal to traverse from input to output by more than one path. However, the mere possession of multiple paths does not require that the function be nonminimum phase. The transfer function of a ladder network is always minimum phase because the zeros are introduced by resonances (at complex frequencies) in the series and shunt elements, which are two-terminal networks. A ladder network has a structural form depicted by Fig. 3.9.

Minimum-phase networks have received considerable attention in the literature; the phase function of a minimum-phase network along the $j\omega$ axis has been proved (by other than p-z arguments) to be uniquely specified in terms of the amplitude function, and methods have been derived for finding the phase function knowing only the amplitude function and conversely. By considering p-plane plots such as those of Fig. 3.8, it is not difficult to see that there exists only a finite number of phase functions for a given amplitude function, and that when the network is specified to be minimum phase there can exist but a single phase function for a given amplitude function.

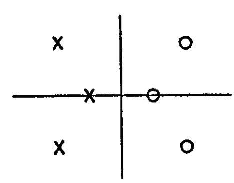


Fig. 3.10. The all-pass function.

Consider the p-z plot of Fig. 3.10. This plot is characteristic of a nonminimum-phase network. However, it is more than this. Notice that each pole in the left half-plane is accompanied by a zero placed as if by mirror reflection in the right half-plane. The reader should have no difficulty in noting that the amplitude function is *constant* along the $j\omega$ axis. Such a function is termed an "all-pass" function because it

provides phase shift with constant amplitude. In general, an all-pass transfer function is defined as a nonminimum-phase function having only poles in the left half-plane and only zeros in the right half-plane so arranged that the zeros are the mirror images of the poles with respect to the $j\omega$ axis.

Physically, all-pass networks are usually constructed as a "bridge" or "lattice." The bridge circuit is shown in Fig. 3.11. The ratio of e_2 to

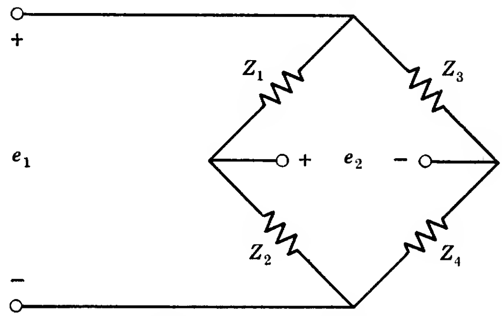


Fig. 3.11. The bridge.

e_1 for this circuit when no source or load impedances exist is easily obtained (using the voltage divider concept) as

$$\frac{e_2}{e_1} = \frac{Z_2}{Z_1 + Z_2} - \frac{Z_4}{Z_3 + Z_4} \tag{3.16}$$

The bridge has long been known to be the most general of all networks. Any stable transfer function, no matter how complex, can always be realized with a lattice, provided (because of incidental dissipation) the poles of the transfer function are not assumed to be too near the $j\omega$ axis.

3.8 Subsidiary network classification

Networks generally perform certain prescribed functions which can be broadly defined on the basis of their p-z locations. First, we shall consider functions which have various amplitude-versus-frequency characteristics. Of course, some networks may be designed to perform several duties simultaneously, but we shall restrict attention to the principal types.

A "low-pass" transfer function passes all frequencies from zero to some "cutoff" frequency and attenuates signals above this upper frequency. Typical p-z plots of such transfer functions are shown in Fig.

3.12. These transfer functions usually yield a phase lag at low frequencies and for this reason are sometimes called "lag" networks.

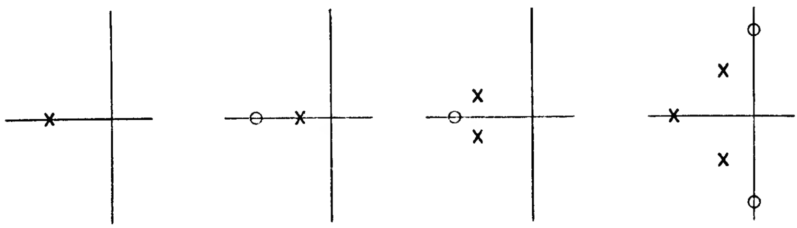


Fig. 3.12. Low-pass functions.

A "high-pass" function passes frequencies above a certain cutoff frequency and attenuates frequencies below this cutoff value. Typical p-z plots of such functions are shown in Fig. 3.13. It is to be noted that

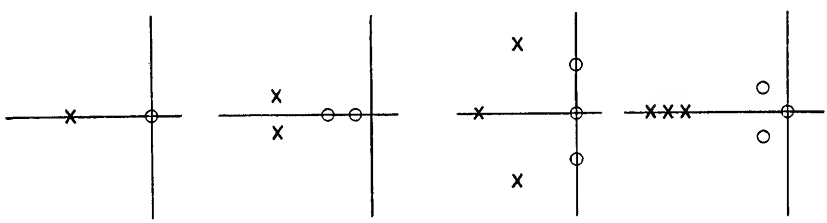


Fig. 3.13. High-pass functions.

high-pass functions usually provide a phase lead at low frequencies and are sometimes called "lead" networks for this reason. Ideally, the number of poles equals the number of zeros.

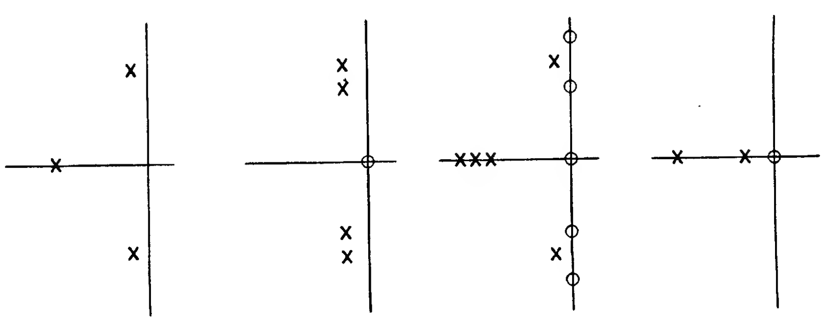


Fig. 3.14. Band-pass functions.

"Band-pass" functions pass only those signals whose frequencies fall between certain finite frequencies. Such functions are exemplified by Fig. 3.14.

"Null functions" (or "band-elimination" functions) reject a certain frequency or band of frequencies. Typical functions are shown in Fig. 3.15. The existence of null networks is the basis of most precision electrical measuring techniques. All a-c measuring bridges are adjusted to give a zero on the $j\omega$ axis so that an input sine wave at the frequency of the zero is attenuated to the greatest possible extent in passing through the network. The adjustment is highly critical, which permits a precise measurement to be performed. Usually, a bridge network is employed,

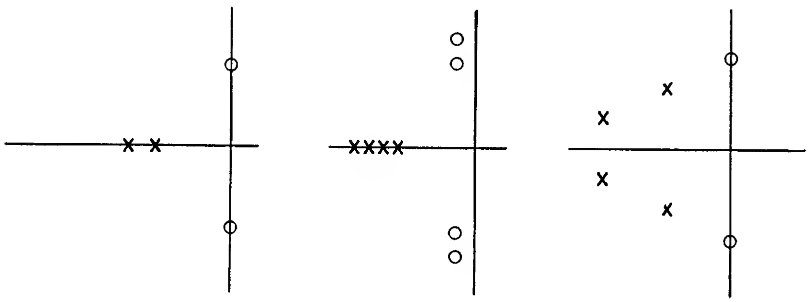


Fig. 3.15. Null functions.

but other networks that provide zeros on the $j\omega$ axis such as the "twin T" are also suitable. Ideally, null functions have equal numbers of p-z, although this is not important in most cases.

Another classification of network functions can be made on the basis of their operational behavior. We originally obtained the complex-frequency variable p as a substitution for the derivative operator which, in the steady state, is the frequency variable $j\omega$. But $j\omega$ was obtained as a result of taking the derivative of the exponential time variable. Hence, if a network transfer function is simply p, the output of the network will be the derivative of the input signal. Similarly, if the network transfer function is $1/p$, the integral of the input will be obtained. Consequently, we can define a derivative network as one containing a zero (or zeros for higher order derivatives than the first) at or very near the origin. It should be noted that high-pass networks tend to behave as derivative networks, particularly if a zero occurs exactly at the origin. Integrating networks show a pole (or poles for higher order integration) near the origin. Low-pass filters tend to behave as integrators, particularly if the bandwidth is very small. In a stable electric network, a pole can never occur at the origin (because of incidental dissipation) but must always be slightly to the left.

A transfer function which delays a time-varying signal but has a unity gain characteristic for all frequencies is not difficult to find. Con-

sider the exponential exp $(j\omega T)$ to operate on a sine wave as

$$\exp(j\omega T)\exp(j\omega t) = \exp[j\omega(t + T)] \tag{3.17}$$

Multiplication by exp $(j\omega T)$ shifts the time axis forward in time by T. It follows that a transfer function having the form exp $(j\omega T) \rightarrow$ exp (pT) advances an input signal by a time T without changing it in any other way. Thus the ideal lead network transfer function is exp (pT) and the ideal lag network transfer function is exp $(-pT)$. If a sinusoidal time-varying signal exp $(j\omega t)$ is looked upon as a phasor, multiplication by exp $(j\omega T)$ advances the phase angle of the phasor by the angle ωT radians, which is a linear angle with frequency. Thus, ideal lead networks have a linear positive phase shift with frequency, while ideal lag networks have a linear negative phase shift with frequency. They both have a magnitude of unity for all frequencies.

The exponential function is not realizable, although it can be approximated. A consideration of it aids in clarifying the mechanism of lead and lag. Most low-pass (lag) networks are at least first-order approximations to the function exp $(-pT)$ at low frequencies, while most high-pass (lead) networks are approximations to exp (pT) at low frequencies.

3.9 Stability of a set of differential equations

In many analyses of linear systems, a set of integrodifferential equations with constant coefficients is obtained. If the set of equations describing (say) a mechanical system can be likened to an electric network, the network becomes an electromechanical analog for the mechanical system in which voltages and currents are the analogs of velocities, positions, and so forth. In order for a stable network to exist, the determinant of the network must have zeros only in the left half-plane, because the zeros of the determinant are the poles of the transfer function. If this is not so, the system of equations does not represent a stable system, either mechanical or electrical, and exponentially increasing voltages or mechanical variables will result.

If the system is stable and the coefficients of the variables are real and constant, an electric analog can be found, perhaps from notions presented in the latter part of Chap. 1. If the network is not bilateral, vacuum tubes must be incorporated in the analog.

3.10 Transient calculations

In proving that the poles of a stable transfer function are limited to the left half-plane, we employed the partial-fraction expansion and solved for the simple transient represented by one of the partial outputs. This general approach can be extended so that the most important

transient calculations can be made in a surprisingly simple and straightforward manner. The methods to be described were originated by Oliver Heaviside, and are applicable to most of the practical transient problems likely to be encountered. The single important exception is where relatively complicated initial conditions exist, although we shall develop methods (in the problems) for handling some initial conditions.

The transient response of a network is also the most general thing to study as it includes every type of response that can be imagined. For example, the steady state is but a part of the response of a network to the unit sinusoid.

Suppose it is desired to find the "impulse response" of a network whose transfer function is $F(p)$, which has simple poles only. The output is therefore

$$e(t) = F(p)E\delta(t) = E\left(C_0 + \frac{A_1}{p - p_1} + \frac{A_2}{p - p_2} + \cdots + \frac{A_n}{p - p_n}\right)\delta(t) \tag{3.18}$$

where $\delta(t)$ is the "unit impulse" function. In eq. 3.18, the usual partial-fraction expansion has been employed, *which reduces the original problem to a number of relatively simple problems.*

The unit impulse is of central importance to the discussion here. Consequently, it is necessary to define it reasonably well. The unit impulse may be considered to be the limiting case of a narrow pulse. The ideal impulse is a pulse that is infinitely narrow and infinitely high, with the width and heighth so proportioned that the area under the pulse is unity. In making a practical measurement of the impulse response of a system, a pulse with a finite width can be used in lieu of the ideal impulse; the approximation is valid as long as the width of the pulse is small compared to the time constants of the network.

The constant C_0 in eq. 3.18 shows that part of the output consists of the impulse itself. Since C_0 is the value of $F(p)$ at $p = \pm j\infty$, the constant C_0 will be zero if the function has more poles than zeros. [If $F(p)$ has one more zero than pole, there will also be a term C_1p in eq. 3.18. However, the constants C_0 and C_1 almost never occur in a practical situation.]

All of the terms of eq. 3.18 are similar, which makes the evaluation of the simple differential equation

$$x(t) = \frac{a}{p + \alpha}\,\delta(t) \tag{3.19}$$

of prime concern. Equation 3.19 can be written in perhaps more

familiar form as in eq. 3.20

$$\frac{dx}{dt} + \alpha x = a\delta(t) \tag{3.20}$$

Let us now interpret $1/p$ as the integral operator and divide both numerator and denominator of eq. 3.19 by p. There results

$$x = \frac{a}{p}\left(\frac{1}{1 + \alpha/p}\right)\delta(t) \tag{3.21}$$

Now expand the denominator of eq. 3.21 into the numerator as

$$x = \frac{a}{p}\left[1 - \frac{\alpha}{p} + \left(\frac{\alpha}{p}\right)^2 - \left(\frac{\alpha}{p}\right)^3 + \cdots\right]\delta(t) \tag{3.22}$$

which makes use of the familiar expansion

$$\frac{1}{1+u} = 1 - u + u^2 - u^3 + u^4 - \cdots \tag{3.23}$$

If the integrations indicated in eq. 3.22 are carried out, we obtain

$$x = a\left(1 - \frac{\alpha t}{1!} + \frac{(\alpha t)^2}{2!} - \frac{(\alpha t)^3}{3!} + \cdots\right) = ae^{-\alpha t} \tag{3.24}$$

which is therefore the solution of the simple differential equation of eq. 3.20. The constant 1 in the series of eq. 3.24 is the first integral of the impulse function; since the area under the unit impulse is unity, its integral is unity. [The integral of the impulse is actually defined as the unit step function $U(t)$, which is a function that rises abruptly from zero to unity at $t = 0$ and is constant at unity for all $t > 0$.]

It is therefore evident that

$$\frac{A_k}{p - p_k} E\delta(t) = A_k E e^{p_k t} \tag{3.25}$$

and hence the solution to eq. 3.18 may easily be determined as

$$e(t) = F(p)E\delta(t) = E[C_0\delta(t) + A_1e^{p_1 t} + A_2e^{p_2 t} + \cdots + A_ne^{p_n t}] \tag{3.26}$$

In finding the solution to the simple differential equation of eq. 3.19, we made use of a series expansion of integrals of the input. This is a rather useful general method which will not be considered further here, although additional data are given in the problems for this chapter. In addition, an example is given in Sec. 11 of Chap. 10.

Let us take the derivative of both sides of eq. 3.25 with respect to the constant parameter p_k, which is an entirely permissible procedure in general. There results

$$\frac{A_k}{(p - p_k)^2}\,\delta(t) = A_k t e^{p_k t} \tag{3.27}$$

If more derivatives are taken and a modicum of mathematical induction employed, the following general equation results

$$\frac{1}{(p - p_k)^n}\,\delta(t) = \frac{t^{n-1}e^{p_k t}}{(n-1)!} \tag{3.28}$$

Now we are able to find the impulse response of transfer functions having multiple-order poles, or *any* function we are likely to encounter. All that need be done is to expand $F(p)\delta(t)$ into partial fractions and write down the answer using eqs. 3.25 and 3.28. For example

$$\begin{aligned}\frac{1}{(p+1)^2(p+2)}\,\delta(t) &= \left(\frac{1}{(p+1)^2} - \frac{1}{p+1} + \frac{1}{p+2}\right)\delta(t)\\ &= te^{-t} - e^{-t} + e^{-2t}\end{aligned} \tag{3.29}$$

Another example, with second-order complex-conjugate poles and a pole at the origin, is

$$\begin{aligned}\frac{1}{p(p^2+2p+2)^2}\,\delta(t) = \Bigg[\frac{A}{p} &+ \frac{B}{[p-(-1+j)]^2}\\ &+ \frac{B^*}{[p-(-1-j)]^2} + \frac{C}{p-(-1+j)} + \frac{C^*}{p-(-1-j)}\Bigg]\delta(t)\end{aligned} \tag{3.30}$$

where the constants are left unevaluated for the purpose of this example. The term $(A/p)\delta(t)$ is simply the integral of $\delta(t)$, which is a constant A (times the unit step function). The solution of eq. 3.30 can be written down directly as

$$\begin{aligned}\frac{1}{p(p^2+2p+2)^2}\,\delta(t) = A &+ Bte^{(-1+j)t} + B^*te^{(-1-j)t}\\ &+ Ce^{(-1+j)t} + C^*e^{(-1-j)t}\end{aligned} \tag{3.31}$$

The application of simple algebra can reduce this to exponentials, sines, and cosines, all having real arguments.

It should particularly be noted that the characteristic transients (that is, the various exponential time dependencies) are determined entirely

by the poles of the function; the zeros affect only the magnitudes and signs of the constants. It is therefore quite evident that for stable systems the transient response is always composed of a constant, simple decaying exponentials, and exponentially decaying sine and cosine waves (except for poles on the $j\omega$ axis, which result in sine and consine waves that do not decay).

A further example (with a zero) is

$$\begin{aligned}\frac{p+1}{(p+2)(p+3)}\,\delta(t) &= (p+1)\left(\frac{1}{p+2}-\frac{1}{p+3}\right)\delta(t)\\ &= (p+1)(e^{-2t}-e^{-3t}) \qquad (3.32)\\ &= -2e^{-2t}+3e^{-3t}+e^{-2t}-e^{-3t} = -e^{-2t}+2e^{-3t}\end{aligned}$$

where the last equation in eqs. 3.32 has been obtained by considering p as the derivative operator d/dt. Thus, zeros can be interpreted as adding derivatives of the solution obtained considering only the poles. If the factor $(p + 1)$ is included in the partial-fraction expansion, eqs. 3.32 become

$$\begin{aligned}\frac{p+1}{(p+2)(p+3)}\,\delta(t) &= \left(-\frac{1}{p+2}+\frac{2}{p+3}\right)\delta(t)\\ &= -e^{-2t}+2e^{-3t}\end{aligned} \qquad (3.33)$$

which is the same solution as that given with the different method of eqs. 3.32. (The method described by eqs. 3.33 is the one usually employed.)

The interpretation of zeros as adding derivatives must be handled with caution when there is a pole at the origin, for then confusion may result in taking a derivative of a step function, or even worse, multiple derivatives of the impulse function. In this event, a function with a pole at the origin, $F(p)/p$, can be reconstructed to have a pole near the origin instead, as $F(p)/(p + \epsilon)$. When the analysis is complete, ϵ can be set equal to zero. However, this problem does not arise if the partial-fraction expansion includes the zeros, as in eqs. 3.33.

At this point, the reader is advised that he is equipped to find the impulse response of any linear circuit with time-independent parameters and for zero initial conditions. The next task is to extend the method so that driving functions other than the impulse can be handled. Before proceeding, it can be mentioned that the response of a function to the unit impulse is often called "Green's" function (because the impulse is essentially a point source).

The impulse response is calculated from the transfer function as $F(p)\delta(t)$. Suppose that we are interested in calculating the response to a general input function $e_0(t)U(t)$ instead of the simple impulse. [The inclusion of the step function $U(t)$ merely signifies that the driving function is zero up to $t = 0$. The $U(t)$ will be ignored henceforth, although its existence is always implied.] We are therefore led to define a network function $E_0(p)$ whose impulse response is the desired input waveform, as

$$e_0(t) = E_0(p)\delta(t) \tag{3.34}$$

Then if the transfer function is defined as

$$e(t) = F(p)e_0(t) \tag{3.35}$$

we can write the desired output as

$$e(t) = [F(p)E_0(p)]\delta(t) = G(p)\delta(t) \tag{3.36}$$

Provided we can find a suitable function $E_0(p)$, which must be defined by p-z as is the transfer function $F(p)$, the response to the more general input $e_0(t)$ is simply calculated as the impulse response of the function $F(p)E_0(p) = G(p)$.

As an example, suppose we wish to find

$$e(t) = F(p)E(e^{-at} - e^{-bt}) \tag{3.37}$$

which is an input that has the shape of a smooth pulse. $F(p)$ is the network transfer function. By working backwards, we easily find

$$\begin{aligned} e^{-at} - e^{-bt} &= \left(\frac{1}{p+a} - \frac{1}{p+b}\right)\delta(t) \\ &= \frac{b-a}{(p+a)(p+b)}\,\delta(t) \end{aligned} \tag{3.38}$$

and thus the desired solution of eq. 3.37 can be written

$$e(t) = \frac{EF(p)(b-a)}{(p+a)(p+b)}\,\delta(t) \tag{3.39}$$

which is calculated just as described before.

Evidently we have arrived at a rather powerful and sophisticated method for treating input signals of rather arbitrary shape, as long as they can be expressed in terms of sine waves, cosine waves, and exponentials. In essence, we have reduced the general problem to that of calculating an impulse response.

Some additional examples are

$$
\begin{aligned}
e(t) &= F(p)U(t) = \frac{F(p)}{p}\,\delta(t) \\
e(t) &= F(p)t = \frac{F(p)}{p^2}\,\delta(t) \\
e(t) &= F(p)\,\frac{d[\delta(t)]}{dt} = pF(p)\delta(t)
\end{aligned}
\tag{3.40}
$$

The first of eqs. 3.40 shows how the important step function is handled; the step function is nothing more than the integral of the impulse. To find the step-function response of $F(p)$, it is only necessary to find the impulse response of $F(p)/p$. If $F(p)$ has no zeros at the origin, the partial-fraction expansion of $F(p)/p$ will have a term $(A/p)\delta(t) = AU(t)$, which is the step function itself. If $F(p)$ has a zero at the origin, the factor $1/p$ will be cancelled and this term will not appear.

The step function is the most significant type of transient input signal to study in connection with low-pass systems. It can further be utilized to study the envelope response of many band-pass systems. In addition, and for linear systems with time-independent parameters, an arbitrary input can be approximated to an arbitrary accuracy with a number of positive and negative step functions occurring at various times after $t = 0$; superposition of the responses to the individual step functions enables the output waveform to be determined. Further, we can determine the response of a system to the time derivative or integral of the step function by finding the derivative or integral of the step-function response respectively. The integral of the step function is the semi-infinite slope function t, whereas the derivative of the step function is the impulse.

Consider now the impulse response of a function having a pair of poles on the $j\omega$ axis. We get

$$
\begin{aligned}
\frac{1}{p^2+\omega^2}\,\delta(t) &= \left(\frac{\frac{1}{2}j\omega}{p-j\omega} - \frac{\frac{1}{2}j\omega}{p+j\omega}\right)\delta(t) \\
&= \frac{1}{2j\omega}\,(e^{j\omega t} - e^{-j\omega t}) = \frac{\sin \omega t}{\omega}
\end{aligned}
\tag{3.41}
$$

which is rewritten as

$$
\frac{\omega}{p^2+\omega^2}\,\delta(t) = \sin \omega t \tag{3.42}
$$

Similarly

$$\frac{p}{p^2 + \omega^2}\,\delta(t) = \cos \omega t \tag{3.43}$$

Equations 3.42 and 3.43 are the important unit sinusoid and cosinusoid respectively. They are manipulated just as before, for example

$$F(p) \sin \omega t = F(p)\,\frac{\omega}{p^2 + \omega^2}\,\delta(t) \tag{3.44}$$

We have succeeded in building up a rather thorough technique for the problems likely to be encountered in circuit theory. We have not, however, derived a correspondingly general method for introducing initial conditions, although if the initial conditions are not too complex (as rarely occurs in a practical situation), they may also be included. The reader should consult the problems for this chapter to see how certain initial conditions can be introduced.

A table of relatively simple functions along with the corresponding impulse responses is a much-used aid. An abbreviated table of this sort is given as Table 3.1. This can alternately be called a table of Laplace

Table—3.1. A Short Table of Laplace Transform Pairs

Function	Impulse Response
p	$\dfrac{d\delta(t)}{dt}$ (unit doublet)
1	$\delta(t)$ (unit impulse)
$\dfrac{1}{p}$	$U(t)$ (unit step function)
$\dfrac{1}{p+\alpha}$	$e^{-\alpha t}$
$\dfrac{1}{(p+\alpha)^n}$	$\dfrac{t^{n-1}e^{-\alpha t}}{(n-1)!}$ (includes $\alpha = 0$)
$\dfrac{1}{(p+a)(p+b)}$	$\dfrac{e^{-at} - e^{-bt}}{b-a}$ (includes $a = 0$) $a \neq b$
$\dfrac{ab}{p(p+a)(p+b)}$	$1 + \dfrac{be^{-at} - ae^{-bt}}{a-b}$ $a \neq b \neq 0$
$\dfrac{\beta}{p^2+\beta^2}$	$\sin \beta t$
$\dfrac{p}{p^2+\beta^2}$	$\cos \beta t$
$\dfrac{\beta}{(p+\alpha)^2+\beta^2}$	$e^{-\alpha t} \sin \beta t$
$\dfrac{\alpha^2+\beta^2}{p[(p+\alpha)^2+\beta^2]}$	$1 - e^{-\alpha t}\left(\cos \beta t + \dfrac{\alpha}{\beta} \sin \beta t\right)$

transform pairs. A table simplifies many transient calculations. For example, $1/[(p + a)(p + b)]$ appears in Table 3.1, which means that the relevant impulse response can be obtained without the usual partial-fraction expansion. Another helpful tabulated impulse response is

$$\frac{1}{p^2 + 2\alpha p + \alpha^2 + \beta^2}\,\delta(t) = \frac{e^{-\alpha t}\sin\beta t}{\beta} \tag{3.45}$$

which can be used to avoid the complicated algebra associated with combining the exponential response of a complex pole and that of its complex conjugate.

The methods we have used to derive the generalized transient response have resulted in a technique indistinguishable from that obtained with the Laplace transform method (or course, within the restrictions imposed). Because we have arrived at the same results, and because the Laplace method is mathematically rigorous, there is no question as to the validity of the results here (although caution must be exercised in extending the Heaviside approach beyond that derived here). The Laplace theory is really no better or easier than the Heaviside method for most transient calculations for systems described by linear integro-differential equations with constant coefficients, except when relatively complicated initial conditions exist; the Laplace method introduces initial conditions in a very straightforward manner.

When the circuit has time-varying parameters and for problems of a more advanced sort, the Heaviside method is not adequate; the Laplace method (or some other mathematical tool) is then required.

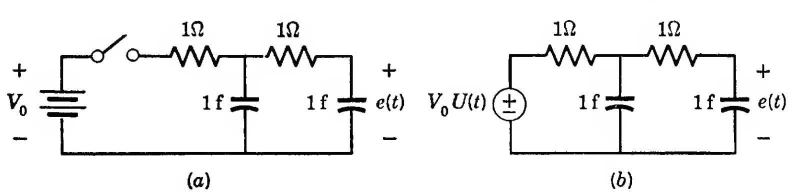


Fig. 3.16. Network for the example.

One physical example may help to clarify the process of finding the transient response. The important steps are enumerated.

The first step is to specify the problem, which is done here with the circuit of Fig. 3.16*a*. At $t = 0$, the switch in Fig. 3.16*a* is closed and the output voltage $e(t)$ is to be determined. The initial voltages on the capacitors are zero. The closure of the switch generates a step-function input; consequently, Fig. 3.16*a* can be redrawn as in Fig. 3.16*b*.

The next step is to find the transfer function. In this, we interpret p as either the derivative operator or the steady-state variable and make use of the node or loop equations. We get

$$e(t) = \frac{1}{p^2 + 3p + 1} e_0(t) = \frac{1}{(p + 2.62)(p + 0.38)} V_0 U(t) \quad (3.46)$$

The third step converts the source into a function of p and an impulse. There results

$$e(t) = \frac{V_0}{p(p + 2.62)(p + 0.38)} \delta(t) \quad (3.47)$$

The fourth step finds the partial-fraction expansion as

$$e(t) = V_0 \left(\frac{1}{p} + \frac{0.17}{p + 2.62} - \frac{1.17}{p + 0.38} \right) \delta(t) \quad (3.48)$$

The fifth step is to write down the transient response

$$e(t) = V_0(1 + 0.17e^{-2.62t} - 1.17e^{-0.38t}) \quad (3.49)$$

which for this example finishes the problem. It may often be necessary to go through a final step in order to put an expression in more readable form; for example, exponentials with complex arguments may be reduced to sines and cosines.

Sometimes, the entire problem may already be solved in a table of Laplace pairs (that is, a table of impulse-response functions). In that case, it is not necessary to do anything more than the first three steps before writing down the answer.

It has been stated that the method developed here is essentially the same as the Laplace method. There is, however, a difference in functional notation. If $F(p)$ is the network transfer function, $e_0(t)$ is the general input, $e(t)$ is the output to be determined, and $E_0(p)$ is a function which gives $e_0(t)$ when excited with the unit impulse, then the Heaviside approach gives $e(t) = F(p)E_0(p)\delta(t)$, where the variable p is interpreted as d/dt. In the Laplace method, all terms in the set of loop or node differential equations are *transformed* by means of an integral transformation, where integration is over time from $t = 0$ to $t = \infty$. Therefore, the Laplace-transformed equations cannot show time as a variable. In Laplace notation, p is the Laplace variable, which is essentially the transformed time variable, $F(p)$ is the Laplace transformation of the network function, and $E_0(p)$ is the Laplace transformation of the driving function. The output must then also be a function of p, $E(p)$. The output is given by $E(p) = F(p)E_0(p)$, which no longer contains

time as a variable. Viewed as a function of p without further interpretation, $F(p)E_0(p)$ as determined by the Laplace method is *identical in all respects* to that obtained by the Heaviside method. The output time function $e(t)$ with the Laplace method is defined as the "inverse" Laplace transformation of the function $E(p)$. Although the inverse can be obtained through complex integration, it is most frequently obtained by expanding $F(p)E_0(p)$ into partial fractions, precisely as in the Heaviside procedure (which is equivalent to the integration).

3.11 Approach to the steady state

The mathematical basis of the all-important steady-state transfer function has in no way been proved up to now; rather, it has only been shown to be "reasonable." Normally, the approach to the steady state is proved with Laplace or Fourier integrals; however, it will be done here in a much simpler manner, which is just as rigorous.

Let a generalized transfer function be $F(p)$, which may be some input or transfer impedance or admittance. The variable p is the differential operator (or Laplace variable, if preferred), and initial conditions are ignored. (In any event, initial conditions are never more than constants, and constants are not important in the steady state.) Let the transient input be the unit sinusoid. Then the transient output is given by

$$e(t) = \frac{\omega F(p)}{p^2 + \omega^2}\,\delta(t) \tag{3.50}$$

The function $F(p)$, which must be a realizable function, is assumed to be absolutely stable; thus all its poles are in the left half-plane (although a simple pole at the origin will be allowed).

The partial-fraction expansion of eq. 3.50 gives

$$e(t) = \left[\begin{pmatrix}\text{Terms for the}\\ \text{poles of } F(p)\end{pmatrix} + \left(\frac{A}{p - j\omega} + \frac{A^*}{p + j\omega}\right)\right]\delta(t) \tag{3.51}$$

At most, the number of zeros in $F(p)$ can exceed the number of poles by unity; therefore the function $F(p)/(p^2 + \omega^2)$ will have *at least* one more pole than zero. Consequently, there will be no constant term in the partial-fraction expansion, nor will there be any terms such as p, p^2, and so forth. If $F(p)$ has a simple pole at the origin, the expansion has a term $1/p$, which gives a partial output that is a constant; however, arbitrary constants are of no interest in the steady state. Except for the possibility of this constant, all the partial outputs related to the poles of $F(p)$ are exponentially decaying. Therefore if we are willing

to wait long enough, we get

$$e(t) \to \left(\frac{A}{p - j\omega} + \frac{A^*}{p + j\omega}\right) \delta(t) \tag{3.52}$$

Now evaluate the constant A

$$A = \left(\frac{\omega F(p)}{p + j\omega}\right)_{p=j\omega} = \frac{F(j\omega)}{2j} \tag{3.53}$$

Equation 3.52 therefore becomes

$$e(t) \to \frac{1}{2j}\left(\frac{F(j\omega)}{p - j\omega} - \frac{F(-j\omega)}{p + j\omega}\right) \delta(t) = \frac{F(j\omega)e^{j\omega t} - F(-j\omega)e^{-j\omega t}}{2j} \tag{3.54}$$

Now $F(j\omega)$ has both a real and an imaginary part as

$$F(j\omega) = R(\omega) + jX(\omega) \tag{3.55}$$

which allows eq. 3.54 to be written

$$\begin{aligned} e(t) &\to (R^2 + X^2)^{1/2} \sin\,[\omega t + \tan^{-1}(X/R)] \\ &= |F(j\omega)| \sin\,[\omega t + \operatorname{Arg} F(j\omega)] \end{aligned} \tag{3.56}$$

and the response for large time is seen to be the original input $\sin \omega t$ with the magnitude and phase of $F(j\omega)$, which is the transfer function $F(p)$ for $p = j\omega$.

The steady state as a limit can also be found when there are poles on the $j\omega$ axis. However, the derivation is considerably more complex. The applied sine wave must be made to grow with time to a constant very slowly. Then the undamped sine waves associated with poles on the $j\omega$ axis will not be excited.

Problems

1. What is the permissible location of the pole of a one-pole function in order that it represent (*a*) a transfer function, and (*b*) an input immittance?
2. A function has one pole and one zero. What are the permissible locations of the pole and the zero in order that the function represent (*a*) a transfer function, and (*b*) an input immittance?
3. A function has two poles. Can it represent an input immittance?
4. A function has one pole and two zeros. Can it represent (*a*) a transfer function, and (*b*) an input immittance?

5. A function has poles at p_1, $p_1^* = -1 \pm j1$ and has a single real zero at $p_2 = -a$. What are the permissible values of a in order that the function represent an input immittance?

6. Repeat Prob. 5 using two zeros and one pole, that is, for the reciprocal of the function.

7. An input immittance has zeros at $p = \pm j$ and poles on the negative real axis. What minimum number of poles is required in order that the function be realizable? What phase-shift restraint is placed upon these poles?

8. Sketch the reactance as a function of frequency for a reactance function having zeros at $p = 0$, $\pm j2$, $\pm j5$, and $\pm j10$ and poles at $p = \pm j1$, $\pm j3$, and $\pm j7$. Show on this single sketch both positive and negative reactance. With a constant multiplier, is this sketch a complete representation of the function?

9. An *R-C* input impedance has poles at $p = -1$, -3, and -5 and zeros at $p = -2$, -4, and -7. This network is used to obtain a transfer function having no zeros. Sketch the amplitude and phase of the transfer function and specify its pole locations.

10. A parallel-resonant circuit has $Q = 10$ and a center frequency of $\omega_0 = 10^6$ radians per second. Plot the magnitude and phase of the input impedance. Also, plot the real and imaginary parts of the input impedance. How do these plots compare with those of the input admittance of a series-resonant circuit having the same Q and center frequency?

11. A parallel-resonant circuit has resistance in the inductance branch and none in the capacitance branch. What is the phase angle at the frequency $1/(LC)^{1/2}$? What must Q be in order that the angle be less than 5 degrees at this frequency?

12. A parallel-resonant circuit has resistance R_1 in the inductance branch and resistance R_2 in the capacitance branch. Plot the p-z of the input impedance. (Assume p-z with imaginary parts.) What relationships must exist between the circuit elements in order that the p-z cancel? In this case, what is the input impedance as a function of frequency?

13. A transfer function $b_0^2/(p^2 + 2ap + b_0^2)$ is defined by two poles. This function can be either of the low-pass or band-pass type. Sketch the amplitude of the transfer function for $a = 2b_0$, b_0, $b_0/2$, and $b_0/4$. In each case, locate the poles. For simplicity, normalize b_0 to unity.

14. What relationship must hold between Z_1, Z_2, Z_3, and Z_4 in order that the transfer function of a bridge be zero at the frequency ω_0? Will this relation be affected by an arbitrary load across the bridge output terminals? Why?

15. A transfer function is $1/(1 + p)$. Compare the amplitude and phase of this transfer function to the exponential e^{-p}. Compare it also to the ideal integrating function $1/p$.

16. The ideal differentiating network has a transfer function of simply p. Why cannot this be realized with a network? Compare the magnitude and phase of the ideal function with the simplest practical differentiating function $ap/(p + a)$. At what frequency would you expect errors to become appreciable using the practical circuit function instead of the ideal function?

17. A positive-real function was defined as one having all its p-z in the left half-plane in complex-conjugate pairs, if complex, and which has a real part that is nonnegative for $p = j\omega$. Show that this pair of requirements is equivalent to the pair of statements

$$\operatorname{Re} Z(p) \geq 0 \quad \text{for} \quad \operatorname{Re} p \geq 0$$

$$Z(p) \text{ real} \quad \text{for} \quad p \text{ real}$$

18. Prove that the coefficients A_j and A_k relating to a pair of complex-conjugate poles in a partial-fraction expansion are complex conjugates.

19. Determine the response e_2 for the circuits of Fig. P.19 when e_0 is the unit step function.

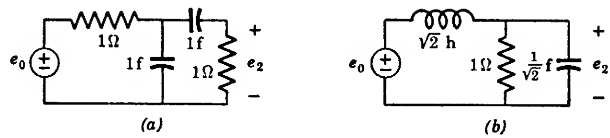


Fig. P.19.

20. Determine the generalized step-function response for the following

$$\frac{1}{(p + a)(p^2 + 2\alpha p + \alpha^2 + \beta^2)}$$

$$\frac{p + a}{p + b} \qquad \frac{p^2}{(p + a)(p + b)}$$

21. The unit step function can be obtained as the limit of the wave of Fig. P.21 as $T \to 0$. By taking the derivative of this waveform, study and discuss the impulse function as a limiting case. In defining the impulse as a limit and in order that the

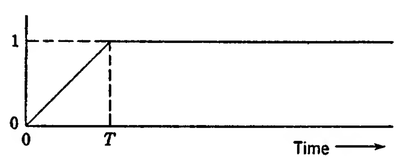


Fig. P.21.

impulse have unit area, is it necessary to restrict the manner in which the step function rises from zero to unity?

22. Let a unit step function at $t = 0$ be applied to a network with the transfer function $H/(p + \alpha)$. At $t = T$, let a negative step function be applied. Sketch the resulting pulse response of the network for $T = 0.5/\alpha$, $1/\alpha$, and $2/\alpha$.

23. A differential equation is

$$\frac{dx}{dt} + kx = A$$

which has the initial condition $x = 0$. A is a constant after $t = 0$. Assume a power series in time as the solution; $x = a_0 + a_1 t + a_2 t^2 + \cdots$. Substitute the assumed solution into the differential equation and equate the coefficients of the powers of time in order to find the solution in terms of well-known functions.

24. From the results of Prob. 23 (by differentiation), find the solution to

$$\frac{dx}{dt} + kx = A\delta(t)$$

25. Find the impulse response of the following

$$\frac{1}{(p + \alpha)(p + \delta)} \qquad \frac{1}{p(p + \alpha)^3}$$

$$\left(\frac{p}{p^2 + 2\alpha p + \alpha^2 + \beta^2}\right)^2$$

26. A capacitor in a system is charged to an initial voltage V_0 as shown in Fig. P.26*a*. By physical reasoning and with the aid of source transformations, show that the system is equivalent for $t \geq 0$ to that of Fig. P.26*b*.

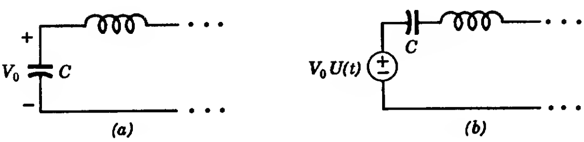


Fig. P.26.

27. An inductor in a system carries an initial current i_0 as shown in Fig. P.27*a*. Show by physical reasoning that the system for $t \geq 0$ is equivalent to Fig. P.27*b*.

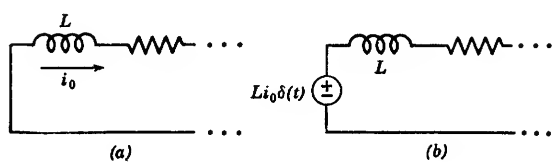


Fig. P.27.

28. Find the impulse response of $p/(p + \alpha)$ from that of $1/(p + \alpha)$ by interpreting p as d/dt.

29. Find the impulse response of $(p + 1)/[p(p + 2)]$ from that of $1/[p(p + 2)]$ by interpreting p as d/dt.

30. Find the impulse response of $p^2/(p + 1)^3$ by interpreting p as d/dt.

31. From the known impulse response of $1/(p + \alpha)$, find by direct integration the step-function response of $1/(p + \alpha)$.

32. A function is

$$\frac{a_0 + a_1 p + \cdots + a_m p^m}{b_0 + b_1 p + \cdots + b_n p^n}\,\delta(t)$$

Divide both numerator and denominator by $b_n p^n$ (assume $m \leq n$) to get

$$\frac{a_m/(b_n p^{n-m}) + \cdots + a_0/(b_n p^n)}{1 + [b_{n-1}/p + \cdots + b_0/(b_n p^n)]}\,\delta(t)$$

Expand in a series of integrals of $\delta(t)$ and from this get the first few terms of the time power series solution.

33. Let $F(p)$ be a transfer function with n poles and m zeros. What is the initial value of the step-function response for $n = m - 1$? $n = m$? $n = m + 1$? $n = m + 2$?

34. A function is

$$e(t) = \frac{1}{(p+1)(p+2)} \delta(t)$$

By means of an expansion of integrals of $\delta(t)$, determine the value of $e(t)$ and its first three derivatives at $t = 0$.

35. Form a "translation" theorem showing that

$$e(t) = F(p)e_0(t - a) = e^{-ap}F(p)e_0(t)$$

where $e_0(t - a)$ is zero for $t < a$ and $e(t)$ is its time-translated counterpart which is zero for $t < 0$. (Use the concept of the ideal lag network.)

36. Find a new impulse response from

$$\frac{\beta}{p^2 + \beta^2} \delta(t) = \sin \beta t$$

by taking the derivative with respect to β.

37. Find a new impulse response from

$$\frac{1}{p + \alpha} \delta(t) = e^{-\alpha t}$$

by taking the indefinite integral with respect to α.

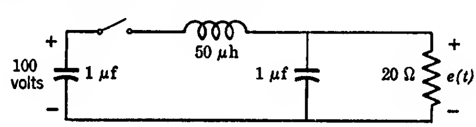


Fig. P.38.

38. Find the output $e(t)$ (numerically) for the system shown in Fig. P.38. The switch is closed on the charged capacitor at $t = 0$.

39. Show that eq. 3.56 follows from eq. 3.54.

4

Elementary Synthesis and Numerical Procedures

In this chapter, we shall introduce some convenient tricks regarding the writing of certain input immittances. This leads into a discussion of the synthesis of L-C, R-C, and R-L networks based upon their pole and zero locations.

In addition, impedance and frequency normalization are discussed in some detail. When it comes to numerical manipulations, *the value of normalization cannot be underestimated.*

Frequently a function must be studied in unfactored form. It often happens that factoring is impractical, in which case some new artifice must be employed to understand the function. Therefore we shall study loci of input and transfer immittances. Because the topic will be taken up later in the book, the treatment in this chapter has been made brief.

Finally, important notions regarding frequency transformations of networks and of functions are discussed. These powerful design tools permit band-pass, high-pass, or band-rejection filter circuits or functions to be obtained from a single low-pass circuit or function. When coupled with normalization, a single low-pass filter normalized to impedance and to frequency can readily be converted to any one of a variety of types of filters having any desired impedance, center frequency, bandwidth, and so forth.

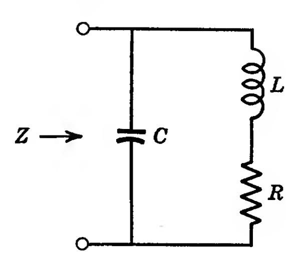


Fig. 4.1. The shunt-peaked circuit.

4.1 A short cut for certain input immittances

Frequently the voltage or current at the input terminals of a bilateral circuit in response to a driving voltage or current is of considerable interest and may represent the answer to a problem. Although the answer may be obtained with a formal determinantal approach, a short cut may often be convenient.

For example, consider the "shunt-peaked" circuit of Fig. 4.1. The input impedance can be found through the following development

$$
\begin{aligned}
Z = \frac{1}{Y} &= \frac{1}{\text{Admittance of } C \text{ branch} + \text{Admittance of } R\text{-}L \text{ branch}} \\
&= \frac{1}{pC + \dfrac{1}{\text{Impedance of } R\text{-}L \text{ branch}}} \\
&= \frac{1}{pC + \dfrac{1}{R + pL}} = \frac{1}{C}\left(\frac{p + R/L}{p^2 + pR/L + 1/LC}\right)
\end{aligned}
\tag{4.1}
$$

Notice that the simple laws of the parallel addition of admittances and series addition of impedances are all that have been used.

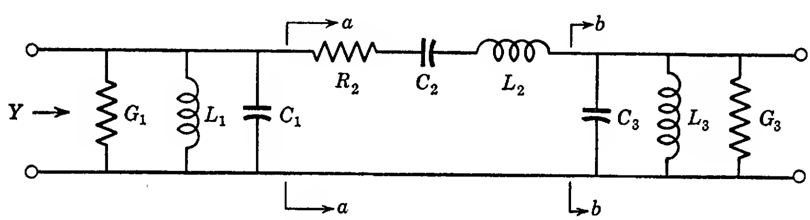


Fig. 4.2. A band-pass filter.

Consider as a more complex example the circuit of Fig. 4.2. In a step-by-step manner, we get

$$
\begin{aligned}
Y &= \text{Admittance of } G_1, C_1, L_1 + \text{Admittance to right of } a\text{–}a \\
&= G_1 + pC_1 + 1/pL_1 + \frac{1}{\text{Impedance to right of } a\text{–}a} \\
&= G_1 + pC_1 + 1/pL_1 \\
&\qquad + \frac{1}{R_2 + pL_2 + 1/pC_2 + \text{Impedance to right of } b\text{–}b} \\
&= G_1 + pC_1 + 1/pL_1 \\
&\qquad + \frac{1}{R_2 + pL_2 + 1/pC_2 + \dfrac{1}{\text{Admittance to right of } b\text{–}b}} \\
&= G_1 + pC_1 + 1/pL_1 + \frac{1}{R_2 + pL_2 + 1/pC_2 + \dfrac{1}{G_3 + pC_3 + 1/pL_3}}
\end{aligned}
\tag{4.2}
$$

The application of simple algebra can reduce this expression to the ratio of two polynomials in p.

The reader has probably noted the similarities between this trick way of writing input immittances and the continued-fraction expansion. In fact, any rational fraction consisting of the ratio of two Hurwitz polynomials (including p-z on the imaginary axis) and having a meaningful continued-fraction expansion can be likened to a specific network expressed as a continued-fraction expansion. This will be taken up again later in this chapter and also in Chap. 6.

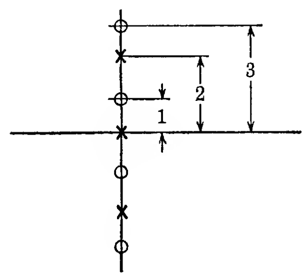


Fig. 4.3. A reactance function.

4.2 Synthesis of input immittances of reactive networks

Suppose a reactance function is given and it is desired to find a corresponding network. The job of finding the network is not at all difficult. The method will be illustrated here by means of the reactance function of Fig. 4.3 and eq. 4.3. (The term reactance function will be generalized to include the susceptance function as well.)

$$F(p) = H\frac{(p^2+1)(p^2+9)}{p(p^2+4)} = H\frac{p^4+10p^2+9}{p^3+4p} \tag{4.3}$$

We may perform a continued-fraction expansion of $F(p)$ in two ways, depending upon whether division is started from the highest or the lowest powers of p, as

$$F(p) = p + \cfrac{1}{p/6 + \cfrac{1}{12p/5 + \cfrac{1}{5p/18}}} \tag{4.4}$$

$$F(p) = \frac{9}{4p} + \cfrac{1}{16/31p + \cfrac{1}{961/60p + \cfrac{1}{15/31p}}} \tag{4.5}$$

where for simplicity the multiplier H has been assumed to be unity.

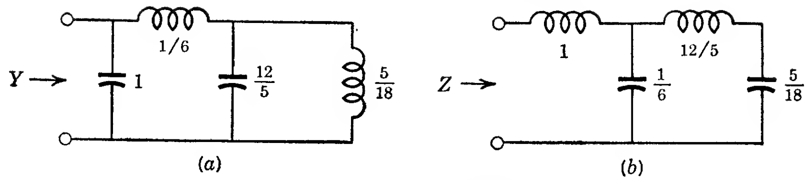


Fig. 4.4. Realizations in ladder form.

Equation 4.4 is realizable as the input admittance and impedance of the networks shown in Fig. 4.4. In addition, eq. 4.5 represents the input impedance and admittance of the networks of Fig. 4.5. In order to include a constant H with a value of other than unity, each term in the numerator of $F(p)$ must be multiplied by this constant before the

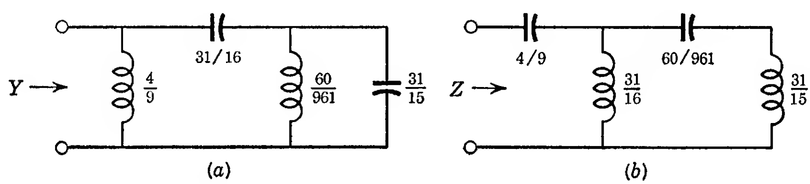


Fig. 4.5. Additional ladder realizations.

expansion is made. (Element values in Figs. 4.4 and 4.5 are in henrys and farads.)

The continued-fraction expansion has given us two realizations of the function as an input impedance and two as an input admittance (two different networks plus their duals). The expansion of the reciprocal of eq. 4.3 does not work because a negative remainder results.

The continued-fraction–expansion method as described comes from Cauer; the resulting networks are said to be Cauer networks. The input immittance of a reactive network can *always* be realized in the form of ladder networks similar to those of Figs. 4.4 and 4.5. In order to keep the Cauer network "pure," the expansion of the original function and each remainder must consistently be done from either the high powers of p or the low powers of p. If some remainders are expanded from the high powers and others from the low powers, a mixed Cauer structure will result, although it may still be a proper network.

The continued-fraction expansion of an impedance (or admittance) may not always be obtainable directly. For example, if the impedance $(p^3 + 2p)/(p^4 + 4p^2 + 3)$ is expanded, negative terms will result, even though it is a perfectly good reactance function. However, if the impedance function is inverted, it becomes an admittance and the con-

tinued-fraction expansion can be made. The resulting network has an admittance that is the reciprocal of the original impedance and hence has the impedance desired from the first. The reason that the expansion of the impedance does not work in this case can be deduced from the behavior of the impedance at $\omega = 0$ and at $\omega = \infty$; the impedance goes to zero at both of these extremes. The first term of the continued-fraction expansion of the impedance starting from the high powers of p is $1/p$, which gives a series capacitor as a first element; however, this cannot be valid if the impedance is to go to zero at $\omega = 0$. Trying division from the low powers of p gives a first term $2p/3$, which represents a series inductor as a first element; however, this cannot be valid if the impedance is to go to zero at $\omega = \infty$. Both Cauer networks result from the expansion of the admittance, starting from both high and low powers of p. One network has a first element that is a shunt capacitor, which is admissible, and the other has a shunt inductor, which is also admissible.

As a second example, consider the impedance $(p^2 + 1)/(p^3 + 2p)$, which goes to zero at $\omega = \infty$ and to infinity at $\omega = 0$. Starting division from the high powers of p, a term $1/p$ is obtained. This gives a series inductor as a first element which is not permissible because the impedance goes to zero at $\omega = \infty$. Starting division from the low powers of p gives a first term $\frac{1}{2}p$, which represents a series capacitor and which is admissible. Thus, one of the Cauer networks is obtained by expanding the impedance from the low powers of p. The other Cauer network must therefore result from a study of the admittance $(p^3 + 2p)/(p^2 + 1)$, which goes to zero at $\omega = 0$ and to infinity at $\omega = \infty$. The first term starting from the high powers of p is p, which gives a shunt capacitor as a leading element and which is admissible. If the expansion of the admittance is started from the low end, the first term is $2p$, which is also admissible; however, the expansion from the low end results in a negative remainder and hence is not permissible. Thus the second Cauer network results from the expansion of the admittance starting with the high powers of p.

It must be observed that the process of synthesis is hardly unique. In fact, we can employ the partial-fraction expansion to find additional network realizations of $F(p)$ (referring to eq. 4.3). The reader is advised to study the special partial-fraction expansions given in the problems for Chap. 2, as they are pertinent to the ensuing discussion.

The partial-fraction expansion of eq. 4.3 is easily found to be

$$F(p) = H\left(p + \frac{\frac{9}{4}}{p} + \frac{15p/4}{p^2 + 4}\right) \tag{4.6}$$

which, for $H = 1$ (in general, each term should be multiplied by H), can be seen to represent one of the immittances of Fig. 4.6 in which elements or pairs of elements are all in series or all in parallel. All that need be done is find the partial-fraction expansion and liken each term to one element or one pair of elements, with each element or pair of elements in a series connection if the function is an impedance, and each element or pair of elements in a parallel connection if the function is an admittance. Two additional but different network realizations can be obtained from the partial-fraction expansion by expanding the reciprocal of $F(p)$. (The reciprocal networks will then be obtained.) The reactance

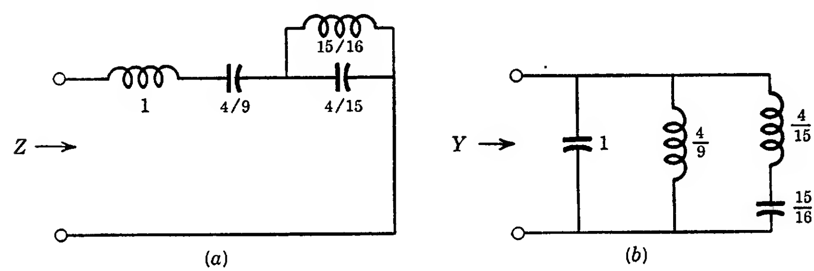


Fig. 4.6. Partial-fraction realizations.

networks arising from the partial-fraction expansion are called Foster networks after their originator. As with realizations derived from the continued-fraction expansion, realizations of input immittances to reactive networks based on the partial-fraction expansion can *always* be found.

With both types of expansions, four different network realizations for any input impedance or input admittance can be obtained. (Including reciprocal networks, there are a total of eight.) These four different networks are called the four "canonic" networks, two of them being Foster networks and two Cauer. (If the function to be realized is very simple, the number of different network configurations may not be as large as four.) *The canonic networks realize a reactance function with the minimum possible number of coils and capacitors.*

Of course, a function may be expanded partly by one method and partly by another. Then, other realizations consisting of mixtures of the basic canonic forms can be obtained.

Cauer networks have one of two basic forms: a ladder with all the L's in series and all the C's in shunt, or a ladder with all the C's in series and all the L's in shunt. These two forms are reciprocal networks.

Foster networks also have two basic forms: a number of parallel-resonant circuits all connected in series with perhaps an extra L and/or

C in series, or a number of series-resonant circuits all connected in parallel with perhaps an extra L and/or C in parallel. These two forms are also reciprocal networks.

A very practical example of L-C synthesis is that where a lumped-circuit approximation to a lossless short-circuited or open-circuited transmission line is obtained. We do not have time to go into transmission-line theory. Nevertheless, we can set down the principles of importance

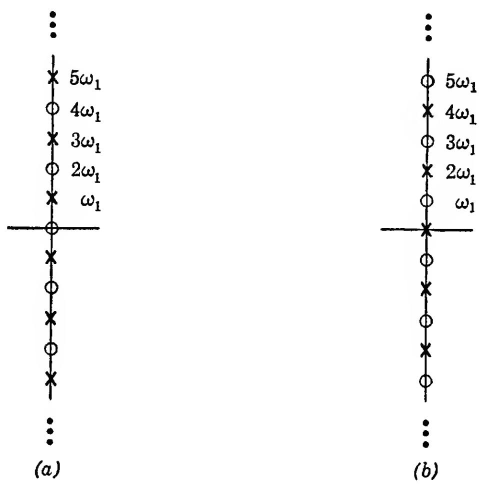


Fig. 4.7. Input impedance of short- and open-circuited transmission lines.

here. A transmission-line equivalent is often found in pulse networks applicable to radar and other systems.

A lossless transmission line of length L meters has an input impedance that is a reactance function of infinite complexity in that it has an infinite number of p-z on the $j\omega$ axis. However, it is in a sense a simple reactance function because all these p-z are equally spaced. If the line is short-circuited at one end, the input impedance has p-z as in Fig. 4.7a, whereas if it is open-circuited, the p-z of the input impedance are as in Fig. 4.7b. In fact, these functions are the same as the tangent function of the exponential or its reciprocal.

For those familiar with transmission-line theory, it will be recalled that the first pole of the impedance of the short-circuited line occurs at a frequence ω_1 such that the line is $\frac{1}{4}$ wavelength long. If v is the velocity of propagation of waves along the line in meters per second and λ_1 is the wavelength in meters corresponding to the frequency ω_1, then $\lambda_1 = 2\pi v/\omega_1$ and $L = \lambda_1/4$, from which $\omega_1 = \pi v/2L$. The input

impedance of the short-circuited line is $Z_0 \tanh(\pi p/2\omega_1)$. The open line has the impedance $Z_0 \coth(\pi p/2\omega_1)$.

An approximation to the open- or short-circuited line valid at least at low frequencies is had with a function containing only the p-z of the actual line that occur at the lower frequencies. As an example (assume for simplicity that $\omega_1 = 1$), let us approximate the input impedance of a short-circuited line with

$$Z_{\text{in}}(p) = \frac{\pi Z_0}{2}\left(\frac{p(1 + p^2/4)}{(1 + p^2)(1 + p^2/9)}\right) \tag{4.7}$$

which is a reactance function which can be realized in the form of any one of the four canonic networks. Z_0 is the characteristic impedance of the line. To be specific, let us find the Foster network for $Z_{\text{in}}(p)$. Normalizing to $Z_0 = 9/8\pi$ for simplicity, we get

$$Z_{\text{in}}(p) = \frac{p^3 + 4p}{(p^2 + 9)(p^2 + 1)} = \frac{Ap}{p^2 + 1} + \frac{Bp}{p^2 + 9} \tag{4.8}$$

which is the same function we used for studying the Foster and Cauer realizations of a reactance function. The synthesis here will obtain the one Foster network we did not get before.

The constants A and B can easily be evaluated to give

$$Z_{\text{in}}(p) = \frac{3p/8}{p^2 + 1} + \frac{5p/8}{p^2 + 9} \tag{4.9}$$

which is the input impedance of the network of Fig. 4.8, and which approximates the input impedance of a short-circuited transmission line up to a frequency of about $3\omega_1 = 3$ radians.

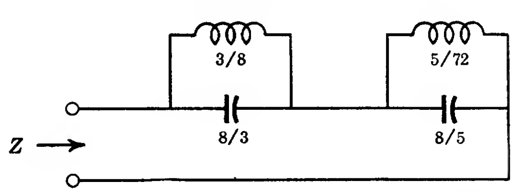


Fig. 4.8. A circuit approximation to a transmission line.

Slightly different approximations (referring to the preceding example) can be had by moving the higher frequency p-z of the approximating function slightly away from the positions of the p-z of the actual transmission line. This can partly account for the infinity of p-z of the actual transmission line that are not present in the lumped network.

We shall not go into this manipulation any further but only mention it here.

4.3 Synthesis of R-L and R-C input immittances

A network containing only R and C or only R and L can be synthesized in a similar manner to that described in the previous section. Analogous canonic forms are obtained. The only difference is that trial and error in forming the continued-fraction expansion may have to be employed.

As an example, consider the function of eq. 4.10

$$F(p) = H\frac{(p+1)(p+3)}{p(p+2)} \tag{4.10}$$

Because of the pole at the origin, eq. 4.10 is restricted to the input impedance of an R-C network or the input admittance of an R-L network. It should be apparent that the reciprocal of eq. 4.10 is restricted to the admittance of an R-C network or the impedance of an R-L network.

The partial-fraction expansion of $F(p)$ is

$$F(p) = H\left(1 + \frac{\frac{3}{2}}{p} + \frac{\frac{1}{2}}{p+2}\right) \tag{4.11}$$

which is given by the networks of Fig. 4.9 if $H = 1$ (where values are in ohms, henrys, and farads).

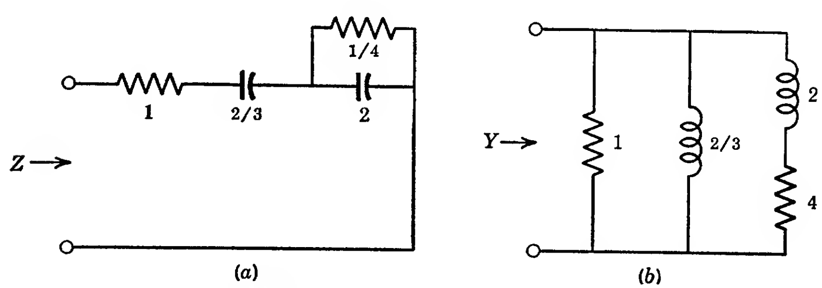


Fig. 4.9. R-C and R-L realizations.

The continued-fraction expansion of eq. 4.10, starting division from the highest powers of p, is

$$F(p) = H\left(1 + \cfrac{1}{p/2 + \cfrac{1}{4 + 6/p}}\right) \tag{4.12}$$

which, for $H = 1$, is given with the networks of Fig. 4.10. Division starting with the lowest powers of p does not work in this case.

Of course, if both continued-fraction and partial-fraction expansions are made of the reciprocal of $F(p)$, some variations may be found. If a mixture of methods is used, a network having a mixed canonical structure may be obtained.

The general canonic form of Fig. 4.9a in which a series connection of simple R-C parallel combinations occurs (along with the associated gen-

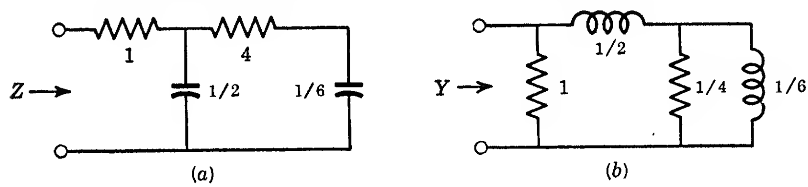


Fig. 4.10. Additional R-C and R-L realizations.

eral partial-fraction expansion), can be studied to advantage in order to understand some of the basic characteristics of R-C input impedances. Let the input impedance be decomposed into its real and imaginary parts evaluated for $p = j\omega$. Then $Z_{\text{in}} = R(\omega) + jX(\omega)$. It is found that $R(\omega)$ is a monotonically decreasing function of frequency which achieves its minimum value at $\omega = \infty$. Also, it is found that the reactance $X(\omega)$ is always negative. The reader should not have too much difficulty demonstrating these properties to his own satisfaction.

4.4 Impedance normalization

Each set of equations describing some arbitrary network has a general form given by one of eqs. 4.13, depending upon whether the equations are written for loop currents or node voltages

$$\begin{aligned}(R_{11} + L_{11}p + S_{11}/p)i_1 - (R_{12} + L_{12}p + S_{12}/p)i_2 \cdots &= e_1 \\ (G_{11} + C_{11}p + \Gamma_{11}/p)e_1 - (G_{12} + C_{12}p + \Gamma_{12}/p)e_2 \cdots &= i_1\end{aligned} \tag{4.13}$$

Let each of the equations of the system be divided by R_{jk} or G_{jk}, whichever is applicable, where R_{jk} or G_{jk} is one of the mutual or self-resistances or conductances respectively. Then, eqs. 4.13 become

$$\begin{aligned}(R_{11}/R_{jk} + L_{11}p/R_{jk} + S_{11}/pR_{jk})i_1 \cdots &= e_1/R_{jk} \\ (G_{11}/G_{jk} + C_{11}p/G_{jk} + \Gamma_{11}/pG_{jk})e_1 \cdots &= i_1/G_{jk}\end{aligned} \tag{4.14}$$

in which case, normalization with respect to resistance or conductance has been made. Because there have been no changes in the derivatives

or integrals of the voltages and currents, normalization has no effect upon the detailed behavior of the system with time or frequency. However, the values of R, L, and C have been changed (as well as amplitudes of voltage and current sine waves) according to

$$\begin{aligned} R_{11} &\to R_{11}/R_{jk} \quad \text{or} \quad G_{11} \to G_{11}/G_{jk} \\ L_{11} &\to L_{11}/R_{jk} \quad \text{or} \quad \Gamma_{11} \to \Gamma_{11}/G_{jk} \qquad (4.15) \\ S_{11} &\to S_{11}/R_{jk} \quad \text{or} \quad C_{11} \to C_{11}/G_{jk} \end{aligned}$$

where the alternative signified by "or" in eqs. 4.15 really amounts to the same thing.

As an example, consider the circuit of Fig. 4.11. Let us normalize this circuit to an impedance level of 1 Ω by causing the normalized value

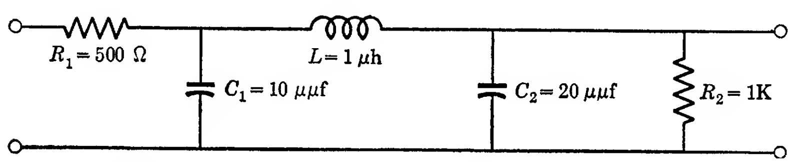


Fig. 4.11. The network before impedance normalization.

of the resistance R_2 to be 1 Ω. To do this, we divide the values of resistance in ohms and inductance in henrys by 1000, and multiply the values of capacitance in farads by 1000. The result is the normalized circuit of Fig. 4.12.

To unnormalize, we go through the reverse procedure, that is, multiply L's and R's and divide C's. We need not, however, go back to the

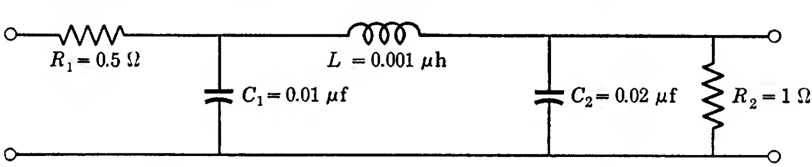


Fig. 4.12. The network after impedance normalization.

original 1000 ohms, but may choose some different value. Thus, the one normalized network can easily be adjusted to fit *any* value of resistance R_2.

A method for obtaining impedance normalization applicable to a transfer- or input-immittance function whose coefficients are expressed in symbols (R, L, and C) rather than numbers is that of setting some

particular R equal to 1 Ω (or some other convenient value). Then, the L's and C's *automatically* assume their normalized values. An example of this procedure will be given in the next section.

A convenience in calculations is given by normalization because a symbol representing some resistance or conductance can be replaced with a pure number (usually unity), thus reducing the number of constants with which it is necessary to deal.

The normalization described in the foregoing is the same as the special case described in Chap. 1 where all rows and columns of a determinant are multiplied by the same constant. Because of its importance, it has been described again here in a somewhat different manner.

4.5 Frequency normalization

Any network input or transfer immittance can be described with a collection of p-z which have fixed distances from the origin of the p plane. With a frequency normalization, we can expand or contract the entire p plane so that the numerical distances of the p-z from the origin are changed without changing their relative positions. Consider, for example, a function containing a zero at $z_1 = -10^6$ radians per second, a zero at the origin, and a pole at $p_1 = -\frac{1}{2} \times 10^6$ radians per second. The function is

$$F(p) = \frac{p(p - z_1)}{p - p_1} = \frac{p(p + 10^6)}{p + 0.5 \times 10^6} \tag{4.16}$$

By substituting a new variable p_x for p related to p by a constant, the nuisance factor 10^6 can be removed. Let us choose $p = 10^6 p_x$. Then

$$F(p_x) = \frac{10^6 p_x(p_x + 1)}{p_x + 0.5} \tag{4.17}$$

which is numerically much simpler to work with. This substitution has contracted the p-z positions by the factor 10^{-6}. The variations of voltage and current in the normalized network function occur at a rate that is 10^{-6} times as fast as that in the unnormalized network.

A slightly different procedure for normalizing to frequency is one in which the product of all the zero positions of a polynomial (except those at the origin, if any) is made equal to unity (along with the proper dimensions). It is normally the polynomial in the denominator of a rational fraction to which this is applied. Of course, if a polynomial exists in the numerator, the same normalization must be used. Consider the general polynomial of eq. 4.18 in which simple term by term division

has put it in a form in which the coefficient of the lowest power of p is unity

$$F(p) = Hp^q(a_m p^m + a_{m-1}p^{m-1} + \cdots + a_2 p^2 + a_1 p + 1) \quad (4.18)$$

Let a new frequency variable be defined by

$$a_m p^m = p_x^{\ m} \quad (4.19)$$

Then, the polynomial takes the form of eq. 4.20 in which the coefficients of both the highest and lowest powers of p are unity; the product of all the zero positions is unity.

$$F(p_x) = H'p_x^{\ q}(p_x^{\ m} + \cdots + 1) \quad (4.20)$$

As a numerical example, the reader can study the normalization described in eq. 4.21

$$F(p) = \frac{12 \times 10^{12}(p + 3 \times 10^6)}{p^3 + 2 \times 10^6 p^2 + 2 \times 10^{12} p + 10^{18}}$$

$$= \frac{12(p/10^6 + 3)}{p^3/10^{18} + (2 \times 10^6 p^2)/10^{18} + (2 \times 10^{12} p)/10^{18} + 1} \quad (4.21)$$

$$F(p_x) = \frac{12(p_x + 3)}{p_x^{\ 3} + 2p_x^{\ 2} + 2p_x + 1}$$

In this example, the product of all the pole positions before normalization is 10^{18}. After normalization, it is simply unity; the poles have been moved 10^{-6} times as close to the origin.

Frequently normalization can be accomplished by normalizing the values of inductance and capacitance in a circuit without reference to equations. The reactance of an inductor is ωL and that of a capacitor is $1/\omega C$. Let us move all the p-z of a network closer to the origin by a factor k, where $k < 1$. In order that the reactances be unchanged at the lower frequencies to which the normalized equations apply, the values of inductance and capacitance must both be increased by the factor $1/k$. For example, consider the shunt-peaked circuit of Fig. 4.13a. We could normalize this circuit from the equations as we did before. However, we can also normalize the circuit directly. Let us

take 10^{-8} for the factor k. Then, the values of both L and C in Fig. 4.13*a* are multiplied by 10^8 to give the normalized circuit of Fig. 4.13*b*. Before or after the frequency normalization, we can effect an impedance normalization. Figure 4.13*c* is that of Fig. 4.13*b* normalized to an impedance level of 1 Ω.

If a transfer or input immittance is expressed in terms of letter symbols (R, L, and C) rather than numerical values, we can normalize to resistance by simply letting some R equal 1 Ω so that all L's and C's (and the rest of the R's) automatically take on their normalized values.

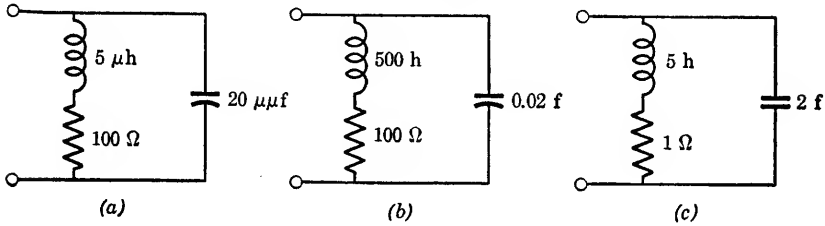


Fig. 4.13. Normalization of the shunt-peaked circuit.

Similarly, we can normalize to frequency by setting some C or L or factor containing C's and L's equal to unity so that normalization to frequency is automatic and the rest of the C's and L's assume their normalized values. For example, consider the transfer function of the series-peaked circuit given in Chap. 2

$$\frac{e_2}{e_0} = \frac{1/(RLC_1C_2)}{p^3 + (1/RC_1)p^2 + [(C_1 + C_2)/(LC_1C_2)]p + 1/(RLC_1C_2)} \tag{4.22}$$

Since only one resistor appears in the circuit, it is evident that the most convenient resistance normalization is given by setting R equal to 1 Ω. Then, eq. 4.22 becomes

$$\frac{e_2}{e_0} = \frac{1/(LC_1C_2)}{p^3 + (1/C_1)p^2 + [(C_1 + C_2)/(LC_1C_2)]p + 1/(LC_1C_2)} \tag{4.23}$$

in which the L's and C's assume their normalized values, not their original ones for R not equal to 1 Ω. It should be noted that we have ignored the dimension of ohms in writing the normalized equations as a matter of simplicity.

Let us normalize to frequency by setting the mean pole position equal to unity. Thus we set $LC_1C_2 = 1$ and eq. 4.23 becomes

$$\frac{e_2}{e_0} = \frac{1}{p^3 + (1/C_1)p^2 + (C_1 + C_2)p + 1} \tag{4.24}$$

in which C_1 and C_2 have values normalized with respect to both impedance and frequency. Equation 4.24 is considerably simpler than the original expression, there being only two parameters. As before, we have ignored dimensions for simplicity.

Alternately, we could effect frequency normalization in eq. 4.23 by assuming L or C_1 or C_2 to be unity. If we take $L = 1$, we get

$$\frac{e_2}{e_0} = \frac{1/(C_1C_2)}{p^3 + (1/C_1)p^2 + [(C_1 + C_2)/(C_1C_2)]p + 1/(C_1C_2)} \tag{4.25}$$

which is a different frequency normalization. As before, only two parameters are involved.

Of course, the equations can be unnormalized with frequency through the reverse procedure. However, it is not necessary to go back to the original p-z positions but rather to any that are desirable as long as they have the same relative positions.

4.6 Graphical plots

When a polynomial having letters or symbols for coefficients is of higher degree in p than two, factoring to find the positions of the zeros is either not possible or not practical. In such an event, little can be done to ease the problem. (If the coefficients are numerical, factoring is always possible.)

Zeros of a polynomial can be either real or complex. If the degree of the polynomial is odd, it has an odd number of zeros, at least one of which must lie on the real axis of the p plane because all zeros, if complex, must occur in complex-conjugate pairs. The number of real zeros must be even (that is, 0, 2, 4, $\cdots$) in an even polynomial. It is well to keep these simple facts in mind because it is generally easiest to use numerical procedures to obtain the real zeros of a polynomial, if any exist, before getting the complex zeros.

In network design as opposed to network analysis, factoring is not always required, as when a network is constructed to fit some already factored rational function of p. However, factoring is sometimes required even in design. The sheer labor of factoring polynomials with numerical coefficients may make it advisable to search for some other device for solving certain problems.

The most illuminating graphical procedure is one in which the magnitude and phase of a function of $p = j\omega$ is plotted as the locus of phasors on a two-dimensional complex-number plane (not the p plane). This new complex plane is called the "locus" plane.

First let us consider the locus of a bilateral input immittance as a function of $j\omega$. The phasors of the function at various frequencies are calculated with ordinary complex algebra and plotted on the locus plane. Then, the line connecting the heads of all these phasors is drawn; it is this line that constitutes the locus. Since the input immittance of all bilateral networks is positive real, the locus must always have a real part that is positive; therefore, the locus must not venture to the left of the imaginary axis. Because the number of p-z can differ at most by unity, the behavior of the locus at very high frequencies can do one of the following:

1. Increase directly with frequency at a phase angle approaching 90 degrees.
2. Decrease as the reciprocal of frequency with a phase angle approaching -90 degrees.
3. Approach a constant at a phase angle approaching zero degrees.

The poles or zeros at the origin must be simple; hence, at very small frequencies the immittance can behave as: 1. $A\omega\underline{/+90°}$; 2. $B/\omega\underline{/-90°}$; or 3. $C\underline{/0°}$, where A, B, and C, are finite real positive constants. Typical locus plots of input immittances are shown in Fig. 4.14. Some impossible loci are indicated in Fig. 4.15. The arrow on these loci indicates the direction of increasing frequency.

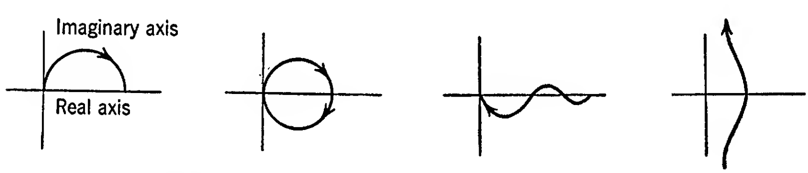


Fig. 4.14. Permissible loci of input immittances.

If vacuum tubes are utilized to obtain an input immittance with a negative real part, the locus will of course venture into the left-hand side of the locus plane. If a transfer function is defined by input impedances when vacuum tubes are operated as pure isolating devices, the individual loci cannot have negative real parts; however, cascading and

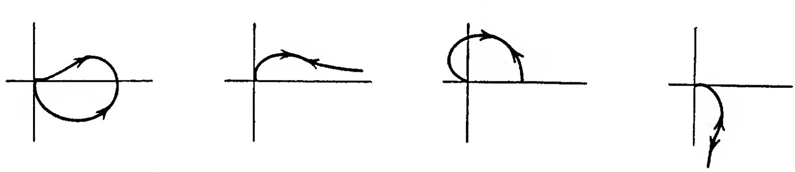


Fig. 4.15. Nonpermissible loci of input immittances.

sign inversion brought about by the tubes can make the over-all locus (which is the product of two or more input immittances) venture any place on the locus plane.

If a zero of the bilateral input immittance occurs on the $j\omega$ axis, the locus must of course go to zero at that frequency and the phase shift must change abruptly by 180 degrees. The only way this can happen without there being a positive real part is for the locus to go through the origin while being tangent to the imaginary axis. If a pole exists on the imaginary axis, the locus must go to infinity at a phase angle of

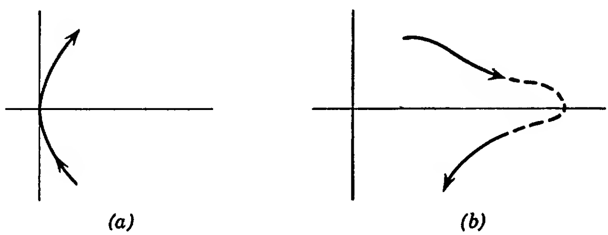


Fig. 4.16. Behavior near a zero and a pole on the imaginary axis.

zero. These two situations are described in Fig. 4.16, in which the arrow directions are not arbitrary. The behavior when the p-z are slightly left of the the $j\omega$ axis is very similar.

A locus plot may also be made of a transfer function. Such a locus is often called a "Nyquist" plot, particularly if it represents a network to which it is planned to add feedback between output and input. The transfer function without feedback is called the "open-loop" transfer function; it is the locus of the open-loop function that is termed the Nyquist plot.

Let us consider the locus plot of a polynomial, which could be the polynomial in either the numerator or (more generally) the denominator of some transfer function. In particular, we shall restrict our interest to Hurwitz polynomials. The various loci appear in Fig. 4.17, where $n = 0, 1, 2, \cdots$ is the number of zeros at the origin. The phase angle forever increases with frequency because of the Hurwitz assumption.

Of most interest is the locus of the Hurwitz polynomial when $n = 0$, that is, when there are no zeros at the origin. If this polynomial represents the denominator of some transfer function, the location of the poles of the function are those of the zeros of the polynomial, and a study of the polynomial tells a great deal about the transfer function, in particular, about how close the poles are to the $j\omega$ axis and whether or not the transfer function is stable.

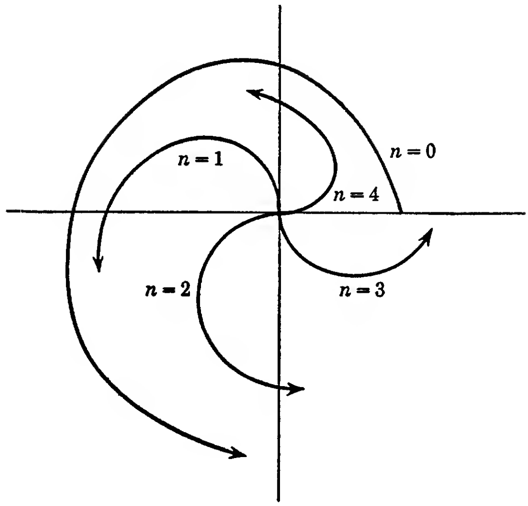


Fig. 4.17. Loci of Hurwitz polynomials.

A polynomial for $p = j\omega$ can be written

$$
\begin{aligned}
F(p) &= a_0 + a_1 p + a_2 p^2 + \cdots + a_n p^n \\
F(j\omega) &= a_0 + a_1(j\omega) + a_2(j\omega)^2 + \cdots + a_n(j\omega)^n
\end{aligned}
\tag{4.26}
$$

The value of $F(j\omega)$ at each $j\omega$ can be found by adding the phasors a_0, $a_1(j\omega)$, $a_2(j\omega)^2$, and so on, in a head to tail fashion in ascending order of

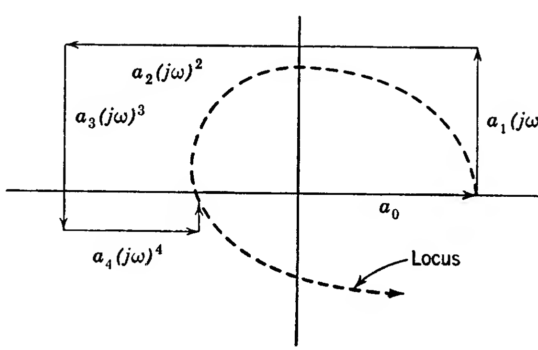


Fig. 4.18. The rectangular plot.

the powers of $j\omega$. Then the plot takes the form of Fig. 4.18. This form of a locus plot for a polynomial can be called a "rectangular" plot.

Once a rectangular plot has been made for one frequency, one for a different frequency is easy to obtain. For example, if the ratio of the two frequencies is 2, the phasor a_0 is unchanged, $a_1(j\omega)$ is twice as long, $a_2(j\omega)^2$ is four times as long, and so on.

If a polynomial not having zeros at the origin is a Hurwitz polynomial, its locus must circle the origin $n/4$ times while increasing continually in phase, and ultimately go to infinity at a phase angle of n times 90 degrees.

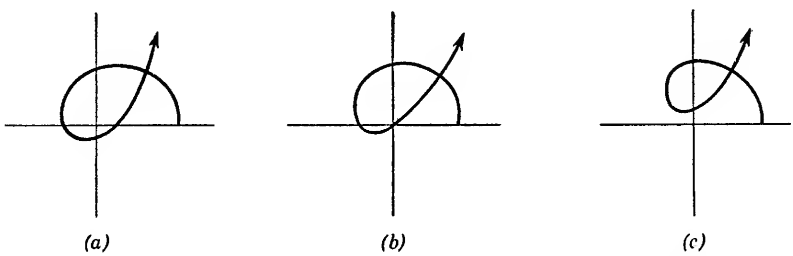


Fig. 4.19. Loci of polynomials having a zero near the $j\omega$ axis.

If not, the polynomial will have zeros on the $j\omega$ axis or in the right half-plane. Consider, for example, the locus of a sixth-order polynomial with a zero: 1. near the $j\omega$ axis in the left half-plane; 2. on the $j\omega$ axis; and 3. near the $j\omega$ axis in the right half-plane, as shown in Figs. 4.19a, b, and c respectively. We could guess at the zero positions from these loci. In particular, those corresponding to Fig. 4.19a would appear roughly as in Fig. 4.20. In order to clarify Fig. 4.19, the reader is advised to set up the corresponding collections of zeros and sketch the phase and amplitude of the polynomial.

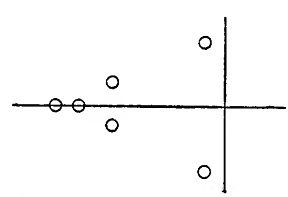


Fig. 4.20. Possible zeros of Fig. 4.19a.

The locus of a transfer function is also often of interest. If the function has a zero near the $j\omega$ axis, the locus will be small at the corresponding frequency, whereas if the function has a pole near the $j\omega$ axis, the function will become quite large. A transfer function with zeros may have an ambiguous phase angle if it is not known whether or not the function is minimum phase. The majority of functions concerned with vacuum-tube circuits are minimum phase; hence, the problem is not too often encountered.

Examples of loci of rational (minimum-phase) fractions are shown in Fig. 4.21. In Fig. 4.21a, there are seven more poles than zeros because

the locus goes to zero at a phase angle of seven times 90 degrees as $\omega \to \infty$. Evidently, no poles or zeros lie very near the $j\omega$ axis. In

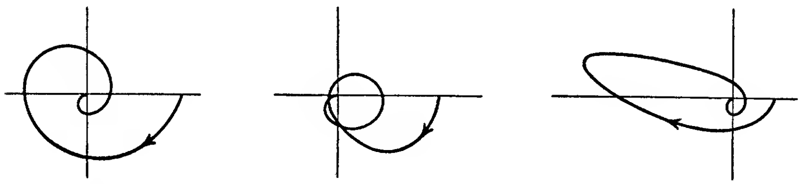


Fig. 4.21. Typical loci of transfer functions.

Fig. 4.21*b* a zero lies near the $j\omega$ axis in the left half-plane, and in Fig. 4.21*c* a pole is near the $j\omega$ axis in the left half-plane. The reader is advised to correlate the loci of Fig. 4.21 and the p-z locations.

4.7 Factorization

The following is a discussion of methods of factoring polynomials having numerical coefficients. Consider an odd polynomial or an even one that is known to contain a real zero. The polynomial is first put in a form (simple division) such that the coefficient of the highest power in p is unity, as in eq. 4.27

$$F(p) = H(p^n + a_{n-1}p^{n-1} + \cdots + a_2p^2 + a_1p + a_0) \tag{4.27}$$

Trial and error in assuming the value of the zero can be made until a factor $p - z_1$ can be divided into eq. 4.27 with a negligible remainder. It is often easiest to employ a special method that converges fairly rapidly if the zeros are not too close together. A trail divisor $p + a_0/a_1$ is used first to obtain

$$\begin{array}{r|l} & p^{n-1} + \cdots \\ \hline p + a_0/a_1 & p^n + \cdots + a_1p + a_0 \\ & \quad\vdots \\ & \qquad\qquad a_1'p + a_0 \\ & \qquad\qquad \underline{a_1'p + a_0'} \\ & \qquad\qquad \text{Remainder} \end{array} \tag{4.28}$$

If the remainder is not negligible, a new factor $p + a_0/a_1'$ is tried

and the process is repeated. Subsequent trials are made until the remainder is negligible.

As an example, consider the polynomial $p^3 + 2p^2 + 3p + 2$. The first trial divisor is $p + \frac{2}{3}$, giving

$$
\begin{array}{r|l}
 & p^2 + (\tfrac{4}{3})p + \tfrac{19}{9} \\
\hline
p + \tfrac{2}{3} & p^3 + 2p^2 + 3p + 2 \\
 & p^3 + (\tfrac{2}{3})p^2 \\
\cline{2-2}
 & \quad (\tfrac{4}{3})p^2 + 3p + 2 \\
 & \quad (\tfrac{4}{3})p^2 + (\tfrac{8}{9})p \\
\cline{2-2}
 & \qquad (\tfrac{19}{9})p + 2 \\
 & \qquad (\tfrac{19}{9})p + \tfrac{38}{27} \\
\cline{2-2}
 & \qquad \text{Remainder}
\end{array}
\tag{4.29}
$$

The remainder is not negligible, so a new trial divisor is taken, $p + 2/(\frac{19}{9}) = p + (\frac{18}{19})$. We get

$$
\begin{array}{r|l}
 & p^2 + (\tfrac{20}{19})p + \tfrac{723}{361} \\
\hline
p + \tfrac{18}{19} & p^3 + 2p^2 + 3p + 2 \\
 & \cdots\cdots\cdots\cdots\cdots\cdots \\
 & (\tfrac{723}{361})p + 2 \\
 & (\tfrac{723}{361})p + 1.897 \\
\cline{2-2}
 & \text{Remainder}
\end{array}
\tag{4.30}
$$

This remainder is also not negligible, so we try $p + 2/(\frac{723}{361}) = p + 0.999$. Division by this factor yields a fairly small remainder. (It is nearly equal to the exact factor $p + 1$ relating to a zero on the negative axis at $p = -1$.)

The number of trials required is largely dependent upon how close together the zeros are. If too close, the process will converge very slowly if at all. Of course, a trial may be made from a guessed zero location, perhaps as obtained from a rectangular plot. Then, convergence may be much more rapid, even though the zeros are close together.

To extract complex zeros, a similar procedure is employed. However, if the zeros are close together as often occurs with band-pass functions, the process may not always be practical.

Let us try to extract a pair of complex zeros from eq. 4.27. The first

trial quadratic factor is taken as $p^2 + (a_1/a_2)p + (a_0/a_2)$ giving

$$\begin{array}{r|l} & p^{n-2} + \cdots \\ \hline p^2 + (a_1/a_2)p + a_0/a_2 & p^n + a_{n-1}p^{n-1} + \cdots + a_2p^2 + a_1p + a_0 \\ & \quad \vdots \\ & \qquad\qquad a_2'p^2 + a_1'p + a_0 \\ & \qquad\qquad \underline{a_2'p^2 + a_1''p + a_0'} \\ & \qquad\qquad \text{Remainder} \end{array} \tag{4.31}$$

If the remainder is not negligible, the trial quadratic is changed to $p^2 + (a_1'/a_2')p + a_0/a_2'$ and the process is repeated.

Another method for extracting complex roots converts the problem to one where real roots can be extracted instead. For example, the polynomial $p^3 + 2p^2 + 2p + 1$ factors as $(p + 1)(p^2 + p + 1)$, which has complex roots at $-\frac{1}{2} \pm j(3)^{1/2}/2$. Let us see if these roots can be obtained through some orderly procedure. If the variable p in the polynomial is changed to $p - \alpha$, and if α is properly chosen, the transformed polynomial will become zero at some $p = \pm j\beta$. The α and β causing the polynomial to become zero are the real and imaginary parts of the root in question. It is advantageous to look at this problem with the aid of the p plane. Substituting $p - \alpha$ for p moves the $j\omega$ axis so that, with α equal to the real part of the zero position, the translated $j\omega$ axis passes precisely through the zero. Our example polynomial therefore becomes $(p - \alpha)^3 + 2(p - \alpha)^2 + 2(p - \alpha) + 1$. If this polynomial is to be zero at some $p = \pm j\beta$, then both its real and imaginary parts must be zero. We therefore have two equations evaluated at $p = \pm j\beta$ as

$$\beta^2(2 - 3\alpha) = (\alpha^3 + 2\alpha^2 - 2\alpha + 1)$$

$$\beta^2 = (3\alpha^2 - 4\alpha + 2)$$

Equating the values of β^2 as given by these two equations results in a cubic in α. Solving this cubic for a *real* α yields the quantity of interest (actually, the real parts of the three zero positions). Substituting this value of α back in either of the two equations for β evaluates the imaginary part of the zero position.

An even more direct procedure than the foregoing for extracting roots is described in the following. The real and imaginary parts of the zero positions are assumed to be α_i and β_i respectively. Then, the polynomial

is equated to its assumed factored form. For example

$$p^4 + ap^3 + bp^2 + cp + d = (p^2 + 2\alpha_1 p + \alpha_1{}^2 + \beta_1{}^2)(p^2 + 2\alpha_2 p + \alpha_2{}^2 + \beta_2{}^2)$$

The coefficients of the powers of p are equated to give a set of simultaneous equations in the unknowns α_1, α_2, β_1, and β_2, and the knowns a, b, c, and d. Substitution or determinantal methods are then used to obtain a set of equations, each containing only one of the unknowns. The α_i and β_i are determined as the real roots of these equations. This method is most convenient when the number of unknown roots is moderate, perhaps three to five.

4.8 Network frequency transformations

The low-pass series-peaked circuit normalized to an impedance level of 1 Ω and having a bandwidth [the frequency where the transfer function falls to $1/(2)^{1/2}$ of that at $\omega = 0$] of 1 radian per second is shown in Fig. 4.22. In a sense, it is a "source" network because it can be converted

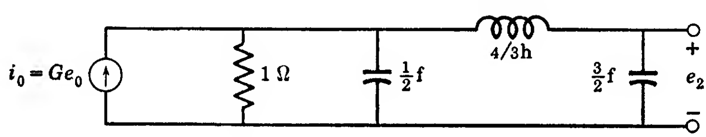


Fig. 4.22. The normalized series-peaked circuit.

to a low-pass network having any desired bandwidth and any desired impedance level by means of impedance and frequency normalizations. It is our purpose here to show how this same simple circuit can be converted to a high-pass, band-pass, or band-elimination network. The one normalized circuit becomes much more general than we could have imagined.

The value of a frequency transformation is very great, particularly that converting a low-pass network to a band-pass network. A band-pass function having good rejection for signals outside the pass band has many more poles and zeros than does its low-pass counterpart. A design procedure that attempts to obtain the band-pass network directly will consequently be far more tedious to apply than one which seeks only to establish a low-pass circuit which can be transformed to a band-pass circuit in a simple manner.

First let us consider the low-pass to high-pass transformation. In this, we want a low-pass network transfer function as depicted by Fig. 4.23 to become a high-pass function such that signals occurring in the pass

band of the low-pass filter will be rejected in the high-pass filter, whereas signals rejected by the low-pass filter will be transmitted by the high-pass filter. In addition, we want a transformation that yields a "crossover" frequency ω_c as shown in Fig. 4.23 about which the transfer functions of the two characteristics are reflected. A transformation that does this replaces the frequency variable ω of the low-pass system with

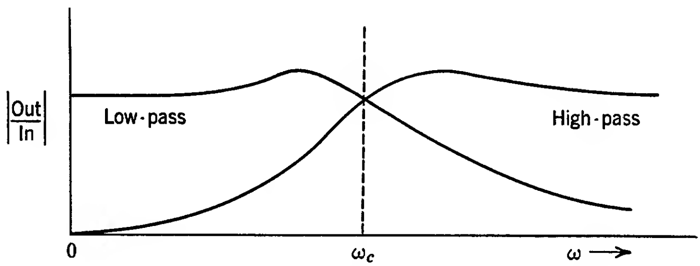


Fig. 4.23. Symmetry of high-pass and low-pass functions.

a new variable ω_c^2/ω'. At $\omega = \omega_c$ and $\omega' = \omega_c$, the two variables ω and ω_c^2/ω' have the same magnitude; hence, the desired crossover frequency is given by this transformation. For $\omega < \omega_c$ in the low-pass filter signals are transmitted, which corresponds to $\omega' > \omega_c$ in the high-pass filter; if $\omega = \omega_c^2/\omega'$ and $\omega < \omega_c$, then $\omega' > \omega_c$.

In a low-pass filter, a capacitor C has a susceptance ωC. Replacing ω by ω_c^2/ω', the susceptance becomes $C\omega_c^2/\omega'$, which is the admittance of an inductor having an inductance of $1/\omega_c^2 C$ henrys. Similarly, the re-

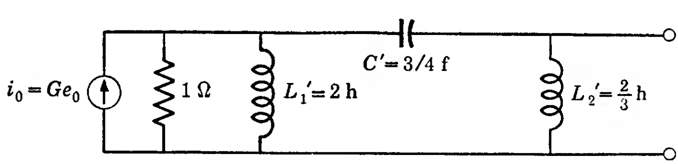


Fig. 4.24. The high-pass filter.

actance of an inductor in a low-pass filter is ωL which, after transformation, becomes $\omega_c^2 L/\omega'$, which is the reactance of a capacitor having a capacitance $1/\omega_c^2 L$.

To sum up, if the upper cutoff frequency of a low-pass filter is designated ω_c (which need not be the half-power frequency) and it is desired to obtain a high-pass filter by means of a frequency transformation having a low-frequency cutoff of ω_c, each capacitor C in the low-pass filter is replaced with an inductor $L' = 1/\omega_c^2 C$, and each inductor L is replaced with a capacitor $C' = 1/\omega_c^2 L$. Resistances are not changed.

The high-pass filter having a low-frequency cutoff of 1 radian (ω_c) derived from the normalized low-pass filter of Fig. 4.22 is shown in Fig. 4.24.

The next transformation to be discussed is the low-pass to band-pass transformation. In this, the frequency variable ω associated with the low-pass system is replaced with the new variable $\omega_0(\omega'/\omega_0 - \omega_0/\omega')$. Then, $\omega' = \omega_0$ in the band-pass system corresponds to $\omega = 0$ in the low-pass system, whereas positive frequencies $\omega > 0$ correspond to $\omega' > \omega_0$ and negative frequencies $\omega < 0$ correspond to $\omega' < \omega_0$. Also, $\omega = \pm B$ in the low-pass function corresponds to

$$\omega_0(\omega'/\omega_0 - \omega_0/\omega') = \pm B \tag{4.32}$$

Solving eq. 4.32 for ω', we get

$$\omega' = \pm(B/2) \pm [(B/2)^2 + \omega_0^2]^{1/2} \tag{4.33}$$

The correspondence can be seen more clearly by observing the amplitude characteristics of Fig. 4.25. For the band-pass function at positive

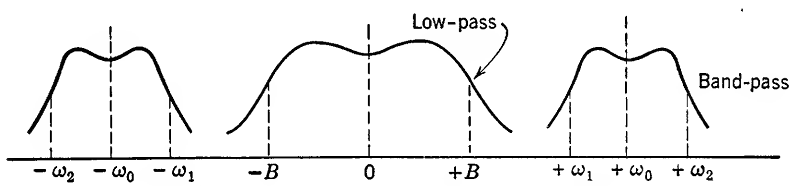


Fig. 4.25. Low-pass and band-pass functions.

frequencies, there are two frequencies corresponding to $\omega = \pm B$. These two frequencies as given by eq. 4.33 are

$$\begin{aligned} \omega_1 &= \omega_0\{[1 + (B/2\omega_0)^2]^{1/2} - B/2\omega_0\} \\ \omega_2 &= \omega_0\{[1 + (B/2\omega_0)^2]^{1/2} + B/2\omega_0\} \end{aligned} \tag{4.34}$$

If we take the geometric mean of these two frequencies, we get

$$(\omega_1\omega_2)^{1/2} = \omega_0 \tag{4.35}$$

which means that about the center frequency ω_0 the band-pass function has geometric rather than arithmetic symmetry such that the half of the band pass below ω_0 will be somewhat more compressed than the half above ω_0; that is, $\omega_0 - \omega_1 < \omega_2 - \omega_0$.

The difference between ω_1 and ω_2 is

$$\omega_2 - \omega_1 = B \tag{4.36}$$

In the low-pass filter, the susceptance of a capacitor is ωC. In the band-pass filter, this becomes

$$j\omega_0 C(\omega'/\omega_0 - \omega_0/\omega') = j\omega' C + \frac{1}{j\omega'(1/\omega_0{}^2 C)} \tag{4.37}$$

which is the admittance of a parallel-resonant circuit with a capacitance C (unchanged) and an inductance $L' = 1/\omega_0{}^2 C$. It can be noted that the resonant frequency of L' and C is the center frequency ω_0. In the low-pass filter, the reactance of an inductor is ωL. Replacing the frequency variable, we get

$$j\omega_0 L(\omega'/\omega_0 - \omega_0/\omega') = j\omega' L + \frac{1}{j\omega'(1/\omega_0{}^2 L)} \tag{4.38}$$

which is the reactance of a series-resonant circuit with an inductance L (unchanged) and a capacitance $C' = 1/\omega_0{}^2 L$. Again, the resonant frequency of L and C' is ω_0.

In summary, suppose that it is desired to design a band-pass filter with a total bandwidth B (measured between the two cutoff points) and a center frequency ω_0. The first step is to design a suitable low-pass filter having a bandwidth B measured from a frequency of zero to the usual upper frequency cutoff point. Then, in shunt with each capacitor in the low-pass filter is placed an inductor of such a value as to resonate with C at the desired center frequency ω_0. In series with each inductor in the low-pass filter is placed a capacitor of such a value as to resonate with L at the center frequency ω_0. The upper and lower cutoff points of the resulting band-pass filter are ω_1 and ω_2 respectively, and $\omega_0 = (\omega_1\omega_2)^{1/2}$ and $B = \omega_2 - \omega_1$. If the "narrow-band" case applies, which occurs when the bandwidth is small compared to the center frequency, then $\omega_0 \cong (\omega_1 + \omega_2)/2$, which is the arithmetic average of the lower and upper cutoff frequencies. Then, the band-pass function will have approximate arithmetic symmetry about the center frequency.

As an example, assume that it is desired to transform the normalized filter of Fig. 4.22 to a band-pass filter having a center frequency of 10 radians and a bandwidth of 2 radians. Since the filter of Fig. 4.22 has a bandwidth of 1 radian, it is first necessary to modify the frequency normalization suitably. To do this, the values of L and C in Fig. 4.22 must be halved to give the circuit of Fig. 4.26, which has the desired bandwidth of 2. Finally, each capacitor is resonated at $\omega_0 = 10$ with an inductor in shunt, and each inductor is resonated at $\omega_0 = 10$ with a capacitor in series, to give the circuit of Fig. 4.27.

In passing, it should be observed that the band-pass equivalent of the simple parallel R-C circuit is the parallel-resonant circuit.

Finally, we shall briefly discuss the low-pass to band-elimination transformation. Rather than go through the details, it should suffice to say that the transformation just described when applied to a high-pass filter instead of a low-pass filter results in a band-elimination filter.

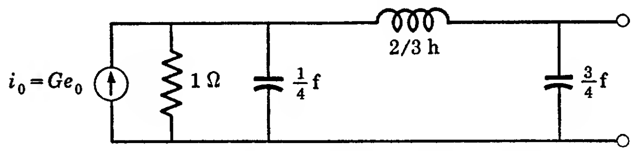


Fig. 4.26. Low-pass filter with increased bandwidth.

If a band-elimination filter having a center frequency ω_0 in the eliminated band of width B is desired, it is first necessary to design a high-pass filter having a low-frequency cutoff of B. Then, *exactly* the same modifications of circuit elements are made as for the band-pass transformation:

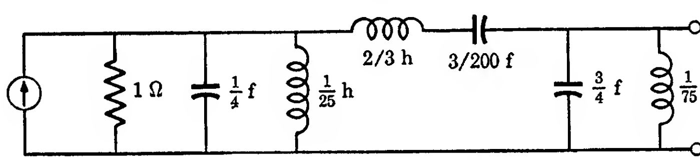


Fig. 4.27. The band-pass filter.

capacitors are placed in series with L's to resonate at ω_0, and inductors are placed in parallel with C's to resonate at ω_0.

4.9 Low-pass to band-pass functional transformations

If the center frequency of a band-pass filter as derived from a low-pass filter is normalized to unity, the low-pass to band-pass transformation is

$$\omega = \omega' - 1/\omega' \tag{4.39}$$

Let us substitute p for $j\omega$ and s for $j\omega'$. Then, this transformation becomes

$$p = s + 1/s = (s^2 + 1)/s \tag{4.40}$$

The variable p can be looked upon as the complex variable related to the low-pass function, whereas s is the complex variable related to the band-pass function. If we let p be the position of a pole of the low-pass function, then we can solve the equation to find the corresponding p-z of the band-pass function. Let p_k be a low-pass pole position. Multi-

plying eq. 4.40 through by $s = s_k$ and solving the quadratic for s_k, we get

$$s_{k1}, s_{k2} = (p_k/2) \pm [(p_k/2)^2 - 1]^{1/2} \tag{4.41}$$

From this, we see that a pole of the low-pass function transforms to two poles in the band-pass function. If p_k is purely real, these two band-pass poles will be complex conjugates. However, if p_k is a pole with an imaginary part, the two band-pass poles will not be complex conjugates. If we stipulate that low-pass poles must be either real or occur as complex conjugates, then the transformation of a low-pass pole

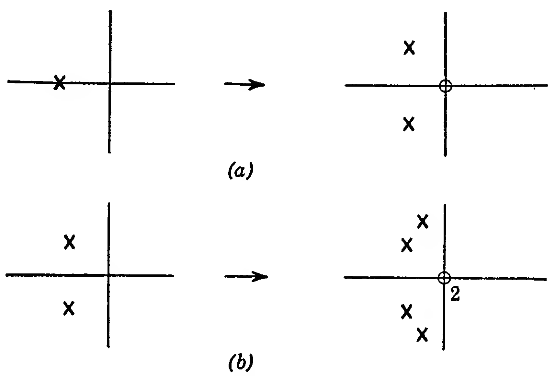


Fig. 4.28. The transformation of poles.

and its complex conjugate will result in *two* pairs of complex-conjugate band-pass poles, and the transformed function will be conjugate analytic and hence obey the rules of network functions.

It will be noted that the transformation of eq. 4.40 has a factor $1/s$ which is lost when multiplying the equation by s in order to obtain a quadratic so that s_k may be solved in terms of p_k. The fact that this factor exists means that the transformation of a low-pass pole not only yields two band-pass poles but also gives a zero at the origin. Thus, a single real pole transforms to a pair of complex-conjugate poles plus a zero at the origin as indicated by Fig. 4.28*a*, whereas a pair of complex-conjugate low-pass poles transforms to two pairs of complex-conjugate band-pass poles plus two zeros at the origin as indicated by Fig. 4.28*b*.

Low-pass zeros transform analogously to low-pass poles; that is, each low-pass zero yields two band-pass zeros plus a pole at the origin.

Let us work through an example to show how a band-pass function can be obtained from a low-pass function by means of the mathematical transformation. Assume that we want a band-pass function that, for simplicity, has a center frequency $\omega_0 = 1$ radian. The bandwidth B

will then usually (but not always) be less than unity. When we finish the filter design, we can obtain whatever center frequency we desire by means of a suitable frequency normalization in which the ratio of ω_0 to B will be unchanged. For example, a center frequency of 10^7 and bandwidth of 10^6, when normalized, become 1.0 and 0.1 respectively. First, we construct a suitable low-pass function with a half-power bandwidth B. Let us take this function to be the three-pole function of eq. 4.42 (where the origin of the proper coefficients will be explained in Chap. 5)

$$\frac{1}{p^3 + 2Bp^2 + 2B^2p + B^3} = \frac{1}{(p + B)(p^2 + Bp + B^2)} \tag{4.42}$$

This transfer function can be realized with the series-peaked circuit. The three poles of the function transform to three pairs of complex-conjugate poles and three zeros at the origin. The single real low-pass pole transforms to two poles according to eq. 4.40 as

$$s_1, s_1{}^* = -(B/2) \pm (B^2/4 - 1)^{1/2} \tag{4.43}$$

The low-pass pole at $p = -B/2 + j(3)^{1/2}B/2$ gives

$$s_2, s_3{}^* = \left(\frac{-(B/2) + [(j(3)^{1/2}B)/2]}{2}\right) \pm \left[\left(\frac{-(B/2) + [(j(3)^{1/2}B)/2]}{2}\right)^2 - 1\right]^{1/2} \tag{4.44}$$

which, for a specific B, can be simplified to numerical real and imaginary parts. The special case $B = 1$ gives

$$s_2, s_3{}^* = \begin{cases} -0.352 + j1.5 \\ -0.148 - j0.63 \end{cases} \tag{4.45}$$

The transformation of the other low-pass pole yields poles that are complex conjugates of s_2 and $s_3{}^*$. For $B = 1$, we get specific values as

$$\begin{aligned} s_1, s_1{}^* &= -0.5 \pm j0.867 \\ s_2, s_2{}^* &= -0.352 \pm j1.5 \\ s_3, s_3{}^* &= -0.148 \pm j0.63 \end{aligned} \tag{4.46}$$

The p-z of the low-pass function and its transformed band-pass function are shown in Fig. 4.29 (for $B = 1$). The center frequency of the band-pass function is unity, which is the geometric mean of the upper and lower cutoff frequencies. The bandwidth between cutoff points is the same as that of the low-pass function (which is unity for this exam-

ple). It should be clear that any value could be used for bandwidth B; unity has been used here only in the interest of simplicity.

The transformation described here corresponds *exactly* to the circuit transformation described in the previous section. Obviously, it is far more convenient to carry out the numerical work from the normalized equations.

A significant simplification takes place if the bandwidth of a band-pass function is considerably less than the center frequency. Then, the collection of low-pass p-z need only be translated along the $j\omega$ axis and

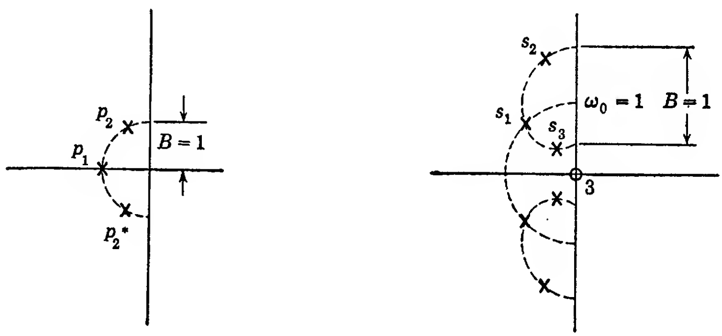


Fig. 4.29. Exact transformation of three low-pass poles.

centered about the desired center frequency. Of course, a complex-conjugate set of p-z must be placed around $-\omega_0$ and the proper number of zeros placed at the origin. However, because bandwidth is measured differently in low-pass and band-pass functions, the p-z of a low-pass function of bandwidth B must have their positions halved before translation if a band-pass function of total bandwidth B is desired.

As an example, assume that a band-pass function of bandwidth B centered at a frequency $\omega_0(\omega_0 \ll B)$ is desired. Let us take the same low-pass function as before, namely, that described by eq. 4.42. The poles of the low-pass function with bandwidth B (measured from $\omega = 0$) are at

$$p_1 = -B$$
$$p_2, p_2^* = -B/2 \pm jB(3)^{1/2}/2 \tag{4.47}$$

Halving these positions, we get

$$p_1' = -B/2 \tag{4.48}$$
$$p_2', p_2'^* = -B/4 \pm jB(3)^{1/2}/4$$

Transforming these sets of poles according to the narrow-band approximation, we get three zeros at the origin plus poles at

$$
\begin{aligned}
s_1, s_1^* &= -B/2 \pm j\omega_0 \\
s_2, s_2^* &= -B/4 \pm j[\omega_0 + B(3)^{1/2}/4] \qquad (4.49)\\
s_3, s_3^* &= -B/4 \pm j[\omega_0 - B(3)^{1/2}/4]
\end{aligned}
$$

This transformation is shown in Fig. 4.30 and gives the pole positions as good approximations to those obtained with the exact analytic expression for B/ω_0 up to about 0.2.

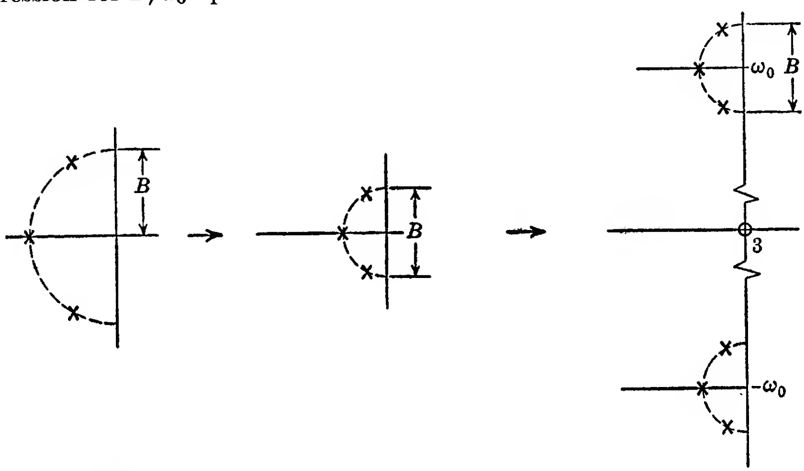


Fig. 4.30. Approximate transformation of three low-pass poles.

The low-pass to band-pass transformation described up to now and in the preceding section gives only one particular class of band-pass functions; that is, it gives the class of band-pass functions derivable from low-pass functions using the particular transformation $s + 1/s$. For each pair of complex-conjugate low-pass poles, the transformation results in two zeros at the origin in addition to the two pairs of transformed complex-conjugate poles. However, many band-pass functions do not have p-z like those described. Most notable are tuned circuits coupled by means of capacitance or inductance. Although these circuits will be taken up in a later chapter, some discussion of the functions themselves is warranted here. Examples are shown in Fig. 4.31. Figure 4.31a is characteristic of the mutually coupled double-tuned circuit. It differs from the transformation discussed up to now because one rather than two zeros occurs at the origin for each two pairs of band-pass complex-conjugate poles. Thus, the mutually coupled double-

tuned circuit function cannot be derived from a low-pass function using the transformation $s + 1/s$. Although a special transformation does exist, it is quite complex and not too important for reasons that will become apparent shortly.

For the narrow-band approximation of Fig. 4.30, the pole placement about ω_0 is similar to that of the low-pass equivalent. This means that, as far as the behavior of the function in this vicinity of frequency ω_0 is

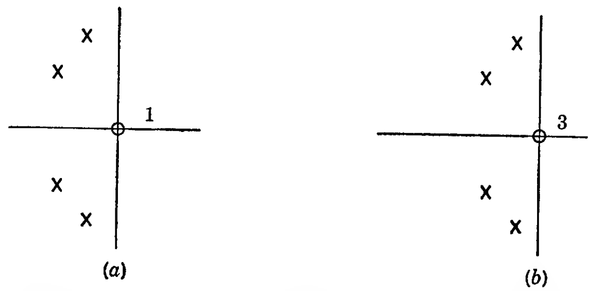


Fig. 4.31. The poles and zeros of double-tuned circuits.

concerned, the effect of the zeros at the origin is essentially cancelled by the effect of the poles grouped around $-\omega_0$. In fact, even if there were only one or two zeros at the origin of Fig. 4.30 rather than three, the phasor lengths from the zeros at the origin and the poles near $-\omega_0$ change so slowly over the frequency range near $+\omega_0$ (which is the restricted frequency range of interest in the narrow-band case) that it

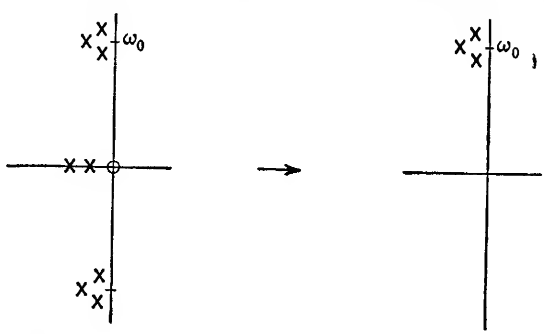


Fig. 4.32. Generalized narrow-band approximation.

makes little difference how many zeros are at the origin. In other words, for any narrow-band band-pass filter function, the effects of all p-z far away from the center frequency ω_0 compared to the bandwidth B can be neglected, that is, treated as a constant (which may, however, be complex). The narrow-band approximation is graphically shown in Fig. 4.32.

From this discussion, it should be clear that a narrow-band transformation can be made to apply to the mutually coupled circuit as well as to others by simply assuming the proper number of zeros at the origin. Whereas the narrow-band approximation of the formal transformation

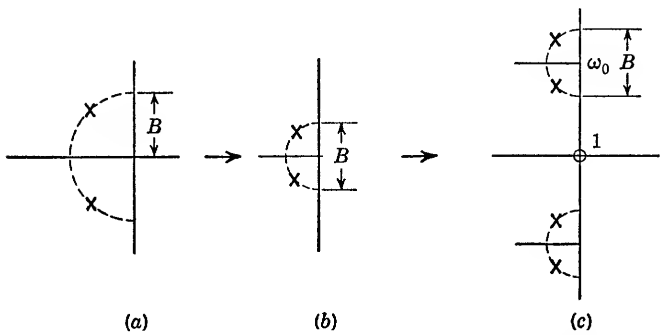


Fig. 4.33. Transformation for the mutually coupled circuit.

$s + 1/s$ is valid for B/ω_0 up to about 0.2, that applicable to the mutually coupled double-tuned circuit having actually fewer zeros at the origin is valid for B/ω_0 as large as unity, which is an unusually large bandwidth-to-center-frequency ratio. Thus, the formal transformation for this case is rarely needed.

As an example, suppose it is desired to design a band-pass function with a center frequency ω_0 and a bandwidth B applicable to the double-tuned mutually coupled circuit. Let us assume that low-pass poles at $-B/(2)^{1/2} \pm jB/(2)^{1/2}$ as in Fig. 4.33a give a low-pass bandwidth of B. Halving these pole positions, we get the configuration of Fig. 4.33b. Finally, going to the band-pass plane and putting one rather than two zeros at the origin, we get the picture of Fig. 4.33c, where there is one zero at the origin and poles at

$$\begin{aligned} s_1, s_1^* &= -\frac{B}{2(2)^{1/2}} \pm j\left(\omega_0 + \frac{B}{2(2)^{1/2}}\right) \\ s_2, s_2^* &= -\frac{B}{2(2)^{1/2}} \pm j\left(\omega_0 - \frac{B}{2(2)^{1/2}}\right) \end{aligned} \tag{4.50}$$

The capacitance-coupled double-tuned circuit function of Fig. 4.31b has three zeros at the origin for each two pairs of complex-conjugate poles. It too can be handled in the narrow-band case. However, the ratio B/ω_0 for this circuit in order that the narrow-band transformation be reasonably accurate is limited to about 0.1.

Problems

1. Use the short cut to determine the input immittances of the networks of Fig. P.1 as ratios of two polynomials in p.

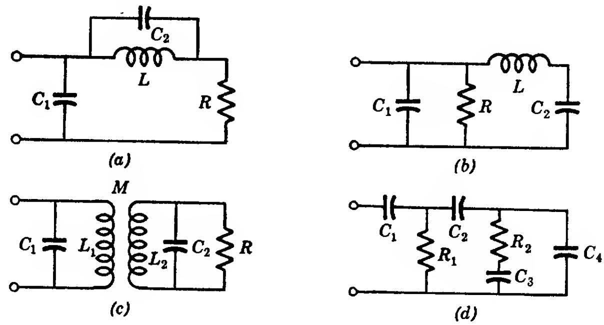


Fig. P.1.

2. The p-z plots of Fig. P.2 are those of the input immittance to reactive networks. For each, find the four canonical networks, all of which have the same input impedance (or admittance).

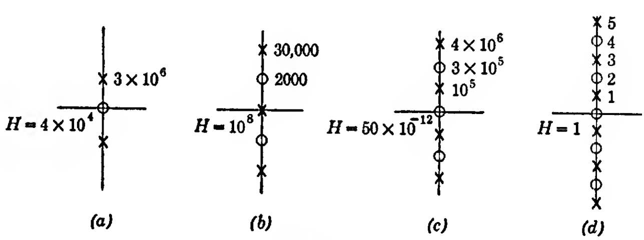


Fig. P.2.

3. Plot the reactance as a function of frequency for the short-circuited transmission line and compare it to the plot of the reactance function used as an example in the text to approximate the short-circuited transmission line.

4. The network represented by the reactance function of Fig. P.2b considered as an impedance is placed in parallel with a 1K resistor. Plot the magnitude of the impedance of the combination as a function of frequency.

5. The p-z plots of Fig. P.5 represent R-C and R-L input immittances. Obtain for each the four canonical networks (where possible) and specify whether the impedance or admittance is synthesized.

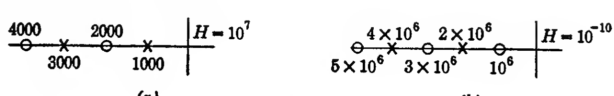


Fig. P.5.

6. The circuit of Fig. P.6 yields a low-pass transfer function that has a bandwidth of 2×10^6 radians per second. Using impedance and frequency normalization, find the corresponding network having a bandwidth of unity and a resistance R of unity.

Then, find the corresponding circuit having a bandwidth of 4 mcs and as high a resistance as is possible assuming that the minimum value of C is 10 $\mu\mu$f.

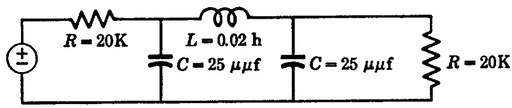


Fig. P.6.

7. Write the transfer function for the circuit of Fig. P.6. Normalize this function to an impedance of 1 Ω and to a mean pole distance of 1 radian. Determine its bandwidth by comparison with the results of Prob. 6.

8. Normalize the following polynomials such that the coefficients of both the zeroth and highest powers of p are unity

$$1 + 20p + 200p^2 + 1000p^3$$

$$1 + 10^{-6}p + 2 \times 10^{-12}p^2 + 5 \times 10^{-18}p^3$$

$$2 + 5 \times 10^{-3}p + 25 \times 10^{-6}p^2 + 10 \times 10^{-9}p^3 + 10^{-12}p^4$$

9. Normalize the polynomials of Prob. 8 so that the coefficients of both the zeroth and first powers of p are unity.

10. Use the results of Prob. 9 to make rectangular plots. From these plots, estimate the locations of the zeros.

11. Determine the loci for the input impedances of the circuits of Fig. P.11.

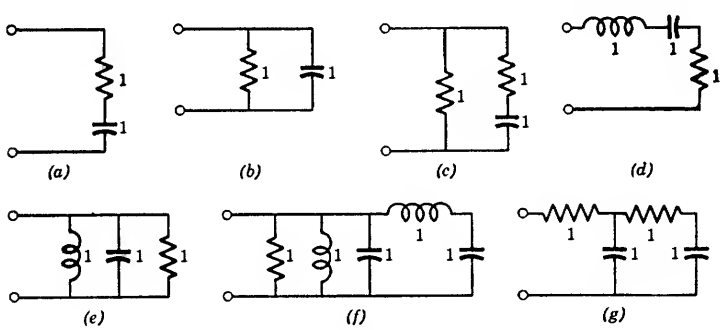


Fig. P.11.

12. Sketch (very qualitatively) the loci associated with the p-z plots of Fig. P.12. Use a sign such that, if the function has a finite value at a frequency of zero, the value is positive.

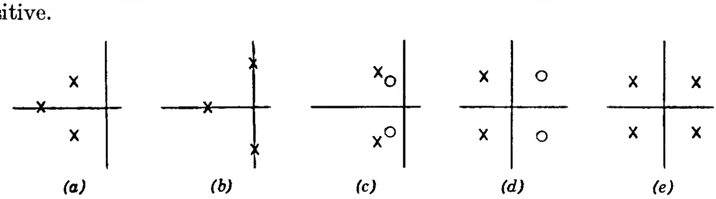


Fig. P.12.

13. Factor the following polynomials. Find the real zero first

$$p^3 + p^2 + p + 1$$
$$p^3 + p^2 + 2p + 1$$
$$p^3 + 2p^2 + p + 1$$
$$p^3 + (2)^{1/2}p^2 + (2)^{1/2}p + 1$$
$$p^3 + 2p^2 + 2p + 1$$
$$p^3 + 3p^2 + 3p + 1$$

14. The polynomial $p^4 + 2p^3 + 10p^2 + 9p + 20$ has two pairs of complex-conjugate zeros. Find them.

15. Convert the low-pass networks of Fig. P.15 to high-pass networks and evaluate the elements in terms of ω_c.

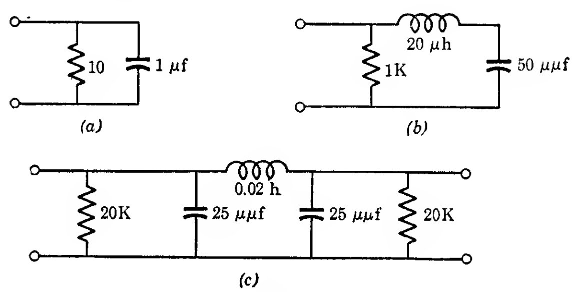


Fig. P.15.

16. Convert the low-pass networks of Fig. P.15 into band-pass networks at a center frequency ω_0. Find element values in terms of ω_0.

17. Use the results of Prob. 15 to construct band-elimination filters with a center frequency ω_0. Evaluate elements in terms of ω_0 and ω_c.

18. A low-pass function of unity bandwidth has no zeros and has poles at $(2)^{-1/2}(-1 \pm j1)$. Use the mathematical transformation to find the pole positions to give a band-pass function having a center frequency 2 and a bandwidth 1. (Use a low-pass bandwidth of $\frac{1}{2}$ and then transform to a center frequency of unity. The resulting pole positions must then be multiplied by 2.)

19. Locate the band-pass poles corresponding to those of Prob. 18 for a band-pass function with a bandwidth of 15 mcs and a center frequency of 30 mcs.

20. Express the four poles of the band-pass function of Prob. 18 in the form of a polynomial in p. Put it in such a form that the coefficient of p^4 is unity.

21. Use the narrow-band approximation to solve Prob. 18 and compare the resulting pole positions with those obtained using the mathematical transformation.

22. Using the narrow-band approximation, find the p-z positions for the double-tuned mutually coupled circuit of bandwidth 10 mcs and center frequency 30 mcs. The corresponding low-pass function of bandwidth 10 mcs has poles at $10(2)^{-1/2}(-1 \pm j1)$ mcs.

23. A constant-k filter (see Chaps. 7 and 8) having a low-pass bandwidth of 2 radians is shown in Fig. P.23. (This and the following problems are intended to build up a practical working knowledge of normalization, transformation, and dual equivalents.)

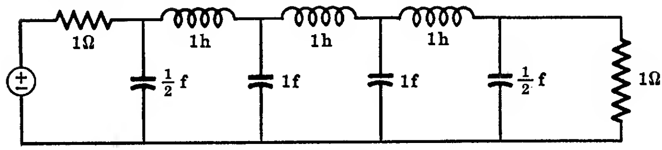


Fig. P.23.

a. Find the filter with a bandwidth of 1 radian and impedance level of 1 Ω. (Work from this in all subsequent problems.)

b. Find the low-pass filter with an impedance level of 600 Ω and a bandwidth of 6 kcs. Do not forget the factor 2π relating a radian and a cycle!

c. Find the low-pass filter with a bandwidth of 4.8 mcs and as high an impedance level as is possible when the minimum possible value of the shunt capacitance at the ends of the filter is 12 μμf.

d. Find the dual of part *b*. (Take all duals with respect to R_0^2, the product of the two terminating resistances.)

24. *a*. Convert the filter of Prob. 23*a* to a high-pass filter with a low-frequency cutoff of 1 radian and an impedance level of 1 Ω.

b. Obtain the high-pass filter with a low-frequency cutoff of 10 kcs and an impedance level of 600 Ω.

c. Obtain the high-pass filter with a low-frequency cutoff of 40 mcs and an impedance level of 300 Ω (which might be used to cut out interference in a television receiver by placing the filter in series with the antenna lead).

d. Find the dual of part *b*.

25. *a*. Convert the filter of Prob. 23*a* to a band-pass filter with a center frequency of unity, an impedance level of unity, and for two bandwidths: 1 radian and 0.1 radian.

b. Obtain the band-pass filter with a center frequency of 20 kcs, an impedance of 1200 Ω, and a bandwidth of 5 kcs.

c. Obtain the band-pass filter with an impedance level of 1200 Ω and band edges at 30 kcs and 60 kcs.

d. Obtain the band-pass filter with a center frequency of 30 mcs, bandwidth of 10 mcs, and as high an impedance level as is possible if the minimum possible value of shunt capacitance at the ends of the filter is 10 μμf.

e. Find the dual of part *b*.

26. Convert the filter of Prob. 23*a* to a band-rejection filter with a rejection band 10 kcs wide, center frequency of 50 kcs, and impedance level of 50K.

27. Convert the low-pass "composite" filter shown in Fig. P.27 to a high-pass filter with a low-frequency cutoff of 1 radian. The low-pass filter has a bandwidth of 1 radian.

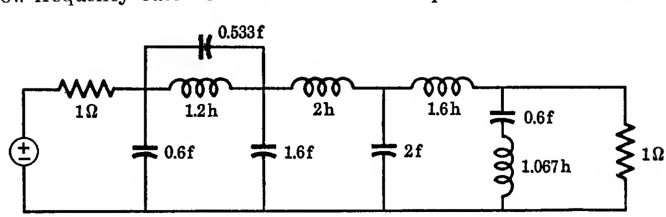


Fig. P.27.

28. Find the dual of the low-pass filter of Prob. 27.

29. Find the dual of the high-pass filter of Prob. 27.

5

Some Important Gain Functions

In Chap. 3 various kinds of functions were classified as low-pass, band-pass, and so on, depending on their pole and zero locations. Other than in a rather general way, the positioning of the p-z in order to realize the desired kind of gain function was not discussed. It is our purpose here to derive and discuss some (but by no means all) very important kinds of functions that yield low-pass characteristics. Of course, the existence of frequency transformations enables us to obtain band-pass and other types of functions from the low-pass function.

In this chapter we shall introduce the "brick-wall" criterion as representing the optimum low-pass function, of course optimum only in the sense of the basic criterion. Then we shall develop various methods for approximating the brick-wall amplitude or phase characteristic with a few p-z to give practical functions that approximate the optimum function.

Finding a network that realizes some desired collection of p-z is called synthesis. Finding a desirable set of p-z is called function design, the subject of this chapter. Clearly, no engineering design can be completed without first obtaining a suitable function.

5.1 Brick-wall and linear-phase functions

The ideal function is one that has a constant gain and a linear phase shift for all frequencies. However, the presence of tube and other capacitances causes the gain and phase characteristics of a practical filter to depart from the ideal. In addition, extraneous signals and noise make it desirable to reduce the gain at high frequencies. If the signals to be passed through the filter have few components above the cutoff frequency, there will be little penalty paid in distortion and the output from the filter will be a fairly faithful reproduction of the input (except for possible magnification due to gain).

What then is the ideal filter function having a finite cutoff frequency? It would appear that the optimum filter should have a constant gain from zero to some finite frequency B and zero gain for all frequencies above B. The phase shift should be linear from $\omega = 0$ to B. Above B it could be anything. Such an ideal filter is specified by Fig. 5.1 and, for obvious reasons, is termed a brick-wall function.

It is apparent that, because of its sharp cutoff characteristics, an *infinite number* of p-z would be required to obtain the ideal brick-wall function. Of course, all practical networks must of necessity be of lim-

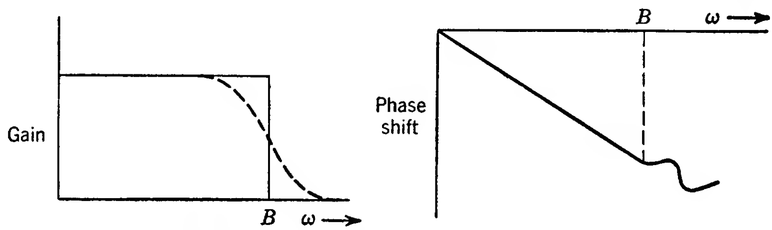


Fig. 5.1. Ideal gain and phase functions.

ited complexity, the inevitable result being that a recourse to approximation is mandatory.

It is difficult to approximate both the brick-wall gain and phase functions simultaneously. However, it is not too difficult to approximate one or the other. This is the method of approach that will be used here. When the ideal gain function is approximated, the phase is ignored. It is "hoped" that the phase will be reasonably linear with frequency. Although it does turn out to be fairly linear, it is not perfectly linear; consequently, other approximation methods may often be more valuable. Conversely, when the ideal phase function is approximated, the gain is ignored. Although the gain will be fairly constant in the pass band, it may not be as constant as is desired.

It must be remembered that a nonideal phase function introduces phase distortion in a signal passing through the network, whereas a nonideal gain function introduces amplitude distortion. It cannot always be said that one kind of distortion is more severe than the other.

The only reasonably simple way to obtain an approximation to both the ideal gain and phase functions is to perform an heuristic sort of average in which the p-z locations for a gain approximation are compared to those of a phase approximation and average p-z positions taken for the function to be used. Often, however, such efforts are not needed, a simpler approximation being quite adequate.

5.2 Taylor approximation to the brick wall

We can approximate any reasonable function by writing a Taylor series about some specific point. If $f(\omega)$ is to be approximated about ω_0, the Taylor series can be written as

$$f(\omega) = f(\omega_0) + \frac{\partial f(\omega_0)}{\partial \omega} \cdot \frac{(\omega - \omega_0)}{1!} + \frac{\partial^2 f(\omega_0)}{\partial \omega^2} \cdot \frac{(\omega - \omega_0)^2}{2!} + \cdots \tag{5.1}$$

If $\omega_0 = 0$, eq. 5.1 is the approximation about zero frequency. The very flat nature of the brick wall makes it apparent that all the derivatives of eq. 5.1 are zero at $\omega = 0$. Thus, eq. 5.1 becomes $f(\omega) = f(0)$ for $\omega < B$.

If some specific gain function agrees with at least part of the expansion of eq. 5.1, it can be considered an approximation in the "Taylor sense" to the brick wall. More precisely, if some specific (power) gain function is expanded about $\omega = 0$, and the first q derivatives of the function are zero at $\omega = 0$, the function is an approximation to the brick wall in the Taylor sense, for example

$$f(\omega) = f(0) + \frac{\partial^q f(0)}{\partial \omega^q} \cdot \frac{\omega^q}{q!} + \text{higher derivatives} \tag{5.2}$$

where the qth derivative is not zero. The resulting function yields a gain characteristic appearing more or less like the dash line in Fig. 5.1. The approximation gets better as the order of the approximation increases, that is, as q increases. In the limit as $q \to \infty$, the approximation tends towards perfection.

We shall concern ourselves with the power gain function. Such functions always have p-z that display quadrantal symmetry; hence, they will consist of the ratio of two even polynomials in p (or ω). All odd derivatives of such functions are zero at $\omega = 0$; hence, the first derivative that has a chance of not being zero at $\omega = 0$ is the second derivative of the function with respect to ω.

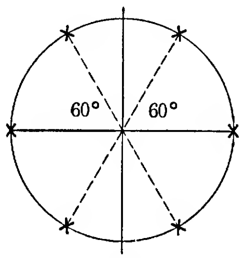
For a normalized bandwidth B of unity, we shall determine the pole locations of the most important Taylor approximation, often called the "maximally flat" or "Butterworth" function. Later in the section we shall show how this specific function of ω is obtained. For the present, it will be arbitrarily presented as

$$f(\omega) = \frac{1}{1 + \omega^{2n}} \to \frac{1}{1 \pm p^{2n}} \tag{5.3}$$

where $\omega = p/j$ and where the sign depends upon whether n is even or odd.

The first $2n - 1$ derivatives of this function at $\omega = 0$ are zero, as can

easily be verified by the reader. Hence, it is a maximally flat function. The polynomial in the denominator is of degree $2n$; consequently, $f(\omega)$ (which is the power ratio) has $2n$ poles which are the roots of the polynomial $1 \pm p^{2n} = 0$; that is, they are the $2n$ roots of ± 1. Note that there is a difference depending upon whether n is even or odd. Example plots are shown for $2n = 6$ and $2n = 8$ in Fig. 5.2.



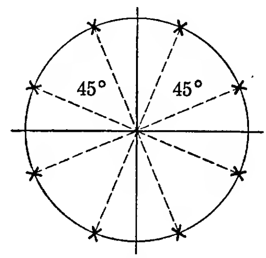


Fig. 5.2. Poles of the maximally flat power gain function.

The $2n$ roots of ± 1 are obtained from a consideration of the exponential. We shall derive the process here in rather general terms. Let y^m be a complex number with a magnitude V and a phase angle θ. Then

$$y^m = Ve^{j(\theta \pm k2\pi)} \tag{5.4}$$

where k is any integer. The reader should recall that a complex number is unchanged by changing its phase angle by any multiple of 360 degrees.

The mth roots of y^m are

$$y = V^{1/m} e^{j(\theta \pm k2\pi)/m} \tag{5.5}$$

which give m distinct values which can be found by allowing k to take on all possible integer values.

In particular, let $V = 1$ and $\theta = 0$ or 180 degrees. Then, eq. 5.4 represents the polynomials

$$y^m - 1 = 0\ (\theta = 0) \qquad y^m + 1 = 0\ (\theta = 180°) \tag{5.6}$$

The m zero positions of these polynomials are found using eq. 5.5 as

$$z_1, z_2, z_3, \cdots = \begin{cases} e^{\pm jk2\pi/m}, & m = 2n, \quad n \text{ odd} \\ e^{j(1 \pm 2k)\pi/m}, & m = 2n, \quad n \text{ even} \end{cases} \tag{5.7}$$

which all lie with *equal angular spacing on the unit circle.*

The function of eq. 5.3 has a value of $\frac{1}{2}$ when $\omega = 1$. Thus, the "half-power bandwidth," or simply the "bandwidth," is unity.

The poles of Fig. 5.2 are those of the power gain; thus they are the poles of the voltage gain plus their negatives. In obtaining the voltage gain function, we are free to delete half the poles of Fig. 5.2. In order to conform to a realizable function, we choose to retain only those poles of the power gain that lie to the left of the $j\omega$ axis. The poles of a maximally flat amplitude function of bandwidth B radians per second lie as shown in Fig. 5.3 for $n = 3$ and 4.

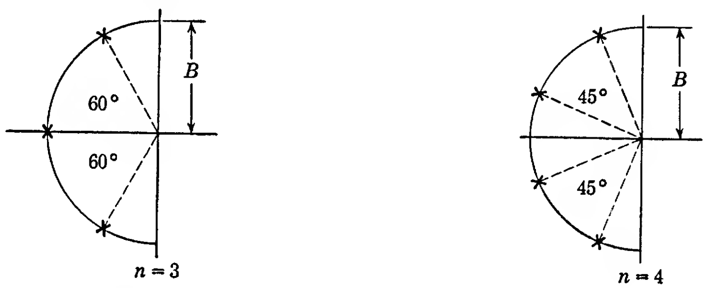


Fig. 5.3. Poles of the maximally flat transfer function.

The function $f(\omega) = 1/(1 - \omega^{2n})$ is also maximally flat. However, this function has poles on the $j\omega$ axis (goes to infinity at $\omega = 1$) and is therefore rejected as an approximation.

In addition to all-pole functions, functions having zeros can be made maximally flat. We shall derive the equations applicable to the general function, which include all-pole functions as a special case.

Consider the generalized transfer power ratio (having quadrantal symmetry) given by

$$VV^* = K\frac{1 + a_2p^2 + a_4p^4 + \cdots + a_{2m}p^{2m}}{1 + b_2p^2 + b_4p^4 + \cdots + b_{2n}p^{2n}} \tag{5.8}$$

where V is the transfer voltage function and VV^* is finite at $\omega = 0$. Let us expand this equation by dividing the denominator into the numerator. Then

$$VV^* = K\left(1 + \frac{(a_2 - b_2)p^2 + (a_4 - b_4)p^4 + \cdots}{1 + b_2p^2 + b_4p^4 + \cdots + b_{2n}p^{2n}}\right) \tag{5.9}$$

A first-order maximally flat function results when the second derivative of the function with respect to ω is zero at $\omega = 0$. This necessitates that

$$a_2 = b_2 \tag{5.10}$$

where we have momentarily interpreted p to be the derivative operator. With this restriction, the lowest power of p in the numerator of eq. 5.9

is p^4. We now observe that a second-order maximally flat system requires that both the second and fourth derivatives of VV^* be zero. Thus, in addition to the requirement of eq. 5.10, we must also have

$$a_4 = b_4 \tag{5.11}$$

in which case the lowest power of p in the numerator of eq. 5.9 is p^6. We can continue in this manner to find that for an Nth order maximally flat system, we must have

$$a_{2j} = b_{2j}, \quad j = 1, 2, 3, \cdots N \tag{5.12}$$

If all the a_{2j} are zero, the b_{2j} must also be zero and we get the all-pole function of eq. 5.3. Clearly, to have an Nth order maximally flat function, the denominator of the expression for VV^* must be of degree at least $2N + 2$.

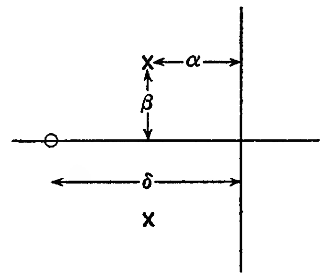


Fig. 5.4. The two-pole, one-zero function.

As a special example, consider the one-zero two-pole voltage transfer function with a p-z plot as shown in Fig. 5.4. We have

$$VV^* = \left(K\frac{p + \delta}{p^2 + 2\alpha p + \alpha^2 + \beta^2}\right)\left(K\frac{-p + \delta}{p^2 - 2\alpha p + \alpha^2 + \beta^2}\right) \tag{5.13}$$

where K is a real constant and where the second term in parenthesis pertains to the p-z in the right half-plane; that is, it contains the negatives of the p-z of the voltage function.

Manipulating eq. 5.13 in order to put it in the form of eq. 5.8, we get

$$VV^* = \frac{K^2\delta^2}{(\alpha^2 + \beta^2)^2}\left[\frac{1 - (1/\delta^2)p^2}{1 + [(2\beta^2 - 2\alpha^2)/(\alpha^2 + \beta^2)^2]p^2 + [1/(\alpha^2 + \beta^2)^2]p^4}\right] \tag{5.14}$$

We may always normalize a function with respect to impedance level. This involves setting some resistance or conductance equal to a fixed number, frequently unity. However, since such impedances do not appear in eq. 5.14, it cannot be done here. It is also possible to normalize to gain. This is usually accomplished by setting K equal to the proper number such that the gain at $\omega = 0$ (at band center) is unity. However, this is of no particular value here. The thing that is of value in simplifying the work is a normalization to frequency. This can be accomplished by setting some pole or zero dimension or some factor

containing a number of pole and zero dimensions equal to some fixed constant, again most frequently unity. We choose to set

$$\alpha^2 + \beta^2 = 1 \tag{5.15}$$

which causes the poles of Fig. 5.4 to lie on the unit circle (which is *not* the bandwidth here). Then, eq. 5.14 becomes

$$VV^* = K^2\delta^2\left(\frac{1 - (1/\delta^2)p^2}{1 + (2 - 4\alpha^2)p^2 + p^4}\right) \tag{5.16}$$

A first-order maximally flat function is obtained by causing eq. 5.10 to be satisfied. Applying this to eq. 5.16, we get

$$\alpha^2 = 1/(4\delta^2) + \tfrac{1}{2} \tag{5.17}$$

Since $a_4 = 0$ and b_4 is not zero, a second-order maximally flat function cannot be realized here.

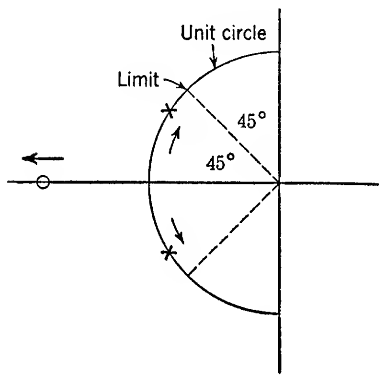


Fig. 5.5. Pole and zero positions for maximal flatness.

The function is not uniquely specified. There are many values of α (or δ) that give the desired maximally flat characteristic. This can probably more easily be seen by observing the way the p-z move while always yielding the maximally flat function as shown in the p-plane sketch of Fig. 5.5. It is of interest to note that as $\delta \to \infty$ (that is, when the zero moves far away from the origin) the function is the maximally flat two-pole function discussed before, which has a bandwidth of unity.

The bandwidth is defined when eq. 5.16 with eq. 5.17 substituted (for $p = j\omega$) is $\frac{1}{2}$. This gives the bandwidth B through

$$\frac{1 + (B^2/\delta^2)}{1 + (B^2/\delta^2) + B^4} = \frac{1}{2} \tag{5.18}$$

from which

$$B^2 = \frac{1}{2\delta^2} + \left[\left(\frac{1}{2\delta^2}\right)^2 + 1\right]^{\frac{1}{2}} \tag{5.19}$$

The function we have given as an example is important enough to warrant the inclusion of a few bandwidth figures. For values of δ (the distance from the origin to the zero) of 0.707, 1, 1.554, 2, 3, 5, 10, and infinity, the bandwidth is 1.55, 1.27, 1.11, 1.06, 1.03, 1.01, 1.005, and 1.00

respectively. The very special case when $\delta = 2\alpha$ ($\delta = 1.554$) corresponds to the shunt-peaked circuit described before. The impedance of the circuit of Fig. 5.6 can realize many different values of $\delta \leq 2\alpha$ with $\delta = 2\alpha$ being the special case when $G_2 = 0$. (For $G_2 = 0$, $G_1 = 1$, and for a bandwidth of 1.11, $L = 0.643$ h and $C = 1.554$ f in Fig. 5.6. These values are readily obtained from the dimensions of the p-z.)

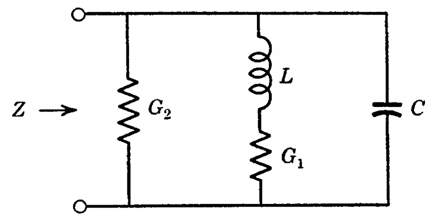


Fig. 5.6. An impedance giving two poles and one zero.

Some further data concerning maximally flat functions are valuable additions here. Let $F(p)$ be an all-pole maximally flat function with unity bandwidth and with $F(0) = 1$. Then

$$F(p)F(-p) = \frac{1}{\pm p^{2n} + 1} \tag{5.20}$$

and we can define

$$F(p) = \frac{1}{p^n + b_{n-1}p^{n-1} + \cdots + b_1 p + 1} \tag{5.21}$$

which has poles only in the left half-plane.

The coefficients b_j can be evaluated in terms of the equal-angle locations of the poles on the unit circle as

$$\begin{aligned} b_{n-1} &= \frac{1}{\sin\theta} \\ b_{n-2} &= \frac{\cos\theta}{\sin 2\theta} b_{n-1} \\ b_{n-s} &= \frac{\cos(s-1)\theta}{\sin s\theta} b_{n-s+1} \end{aligned} \tag{5.22}$$

and the coefficients are symmetric such that $b_{n-1} = b_1$, $b_{n-2} = b_2$, and so forth. The angle $\theta = \pi/2n$.

A general maximally flat function having quadrantal symmetry and with a value of unity at $\omega = 0$ can *always* be written

$$F(p)F(-p) = \frac{1}{1 \pm p^{2n}\{[A(p^2)]/[B(p^2)]\}} \tag{5.23}$$

where the functions A and B in themselves have quadrantal symmetry.

A particularly interesting case is that where A is a constant and B has zeros (in pairs) on the $j\omega$ axis, for example

$$F(p)F(-p) = \frac{1}{1 \pm p^{2n}[A/(p^2 + \omega_1{}^2)^2(p^2 + \omega_2{}^2)^2]} \tag{5.24}$$

When the half-power bandwidth of this function is set equal to unity, and assuming both ω_1 and ω_2 are larger than unity, the constant A can be evaluated to give

$$F(p)F(-p) = \frac{(p^2 + \omega_1{}^2)^2(p^2 + \omega_2{}^2)^2}{(p^2 + \omega_1{}^2)^2(p^2 + \omega_2{}^2)^2 \pm p^{2n}(\omega_1{}^2 - 1)^2(\omega_2{}^2 - 1)^2} \tag{5.25}$$

which is a function having $2n$ poles (which are only roughly located on the unit circle) and four pairs of complex-conjugate zeros, two pairs at $\pm j\omega_1$ and the other two at $\pm j\omega_2$. The larger are ω_1 and ω_2, the closer will be the poles to their positions when there are no zeros, that is, equally spaced on the unit circle.

Of course, the function described in the foregoing can be extended to include any number of zeros on the $j\omega$ axis; two were chosen here only as an illustration.

The value of a function having zeros on the $j\omega$ axis is that the rejection for signals outside the pass band can be made much larger than is possible with a function containing only poles (unless the number of poles is uncomfortably large).

5.3 Bandwidth narrowing

If the transfer function of a circuit is $f(p)$, a cascaded system of N such circuits, each isolated from the rest, has a transfer function $[f(p)]^N$. If N isolated nonidentical vacuum-tube networks having transfer functions $f_1(p)$, $f_2(p)$, $\cdots$ are cascaded, the transfer function for the entire cascaded chain is the product of the individual transfer functions

$$F(p) = f_1(p)f_2(p) \cdots f_N(p) \tag{5.26}$$

In a cascaded system, the first stage drives the second, which in turn drives the third, and so on. By cascading, complicated transfer functions can be constructed from simple and easily interpreted functions, a phenomenon of extreme value.

When identical low-pass functions are cascaded, the over-all system bandwidth is reduced over that of each individual network. For example, the half-power frequency of one network becomes the quarter-power frequency when two such networks are cascaded.

Because the maximally flat all-pole function is widely used, and because its defining equation is simple, the effects of cascading identical

systems can easily be determined. Let N functions, each having the power transfer function $1/(1 + \omega^{2n})$, be cascaded. The half-power bandwidth of each function is obviously unity. The over-all transfer function is

$$F(\omega) = \left(\frac{1}{1 + \omega^{2n}}\right)^N \tag{5.27}$$

and the system half-power bandwidth occurs when $F(\omega)$ is $\frac{1}{2}$. Letting the corresponding frequency be B_N, we get

$$\left(\frac{1}{1 + B_N{}^{2n}}\right)^N = \frac{1}{2} \tag{5.28}$$

from which B_N may be obtained. To be somewhat more general, we can let B be the half-power bandwidth of each individual circuit. Then we can obtain the over-all bandwidth relative to the bandwidth of one circuit as

$$\frac{B_N}{B} = (2^{1/N} - 1)^{1/2n} \tag{5.29}$$

which is known as a "bandwidth-narrowing" equation. This expression includes maximally flat all-pole functions of all orders, including the simple one-pole voltage gain function realizable with a single resistor and a single capacitor.

It can be observed that the system bandwidth decreases rapidly with the number of cascaded functions N but less severely as the order of the maximally flat approximation of each function is increased.

5.4 The Chebyshev approximation to the brick wall

Consider what would happen if all the poles of an all-pole maximally flat function were moved to the right by multiplying their real parts by the same constant k, where $k < 1$, without changing their imaginary parts. This is indicated in Fig. 5.7. The poles now lie on an ellipse. Rather than being flat as is the case when the poles lie on a circle, we might expect that the gain function ripples somewhat and the ripples grow in magnitude as k is decreased. This is indeed the case. In fact, the function ripples evenly and has the same half-power bandwidth as

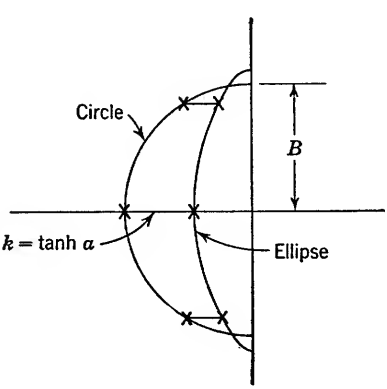


Fig. 5.7. Pole locations of a Chebyshev function.

the maximally flat function. The various points of interest are shown in Fig. 5.8 for a three-pole transfer function.

The poles of a maximally flat function are moved in by the factor $k = \tanh(a)$ (hyperbolic tangent) to give ripples in voltage gain be-

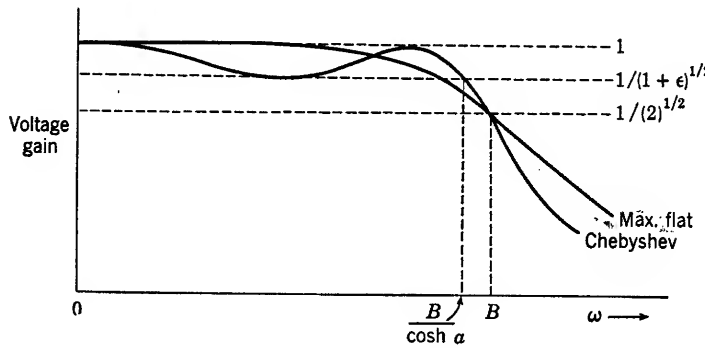


Fig. 5.8. Maximally flat and Chebyshev gain functions.

tween 1 and $1/(1 + \epsilon)^{1/2}$ and the same half-power bandwidth. The parameter a is given by

$$a = (1/n) \sinh^{-1} (1/\epsilon)^{1/2} \tag{5.30}$$

where n is the number of poles of the voltage transfer function (three in Fig. 5.8). The resulting response function is said to be an "equal-ripple" or a "Chebyshev" (Tschebysheff is an alternate spelling) type of function, the name arising from the presence of a Chebyshev polynomial in the expression for the power gain. For the same bandwidth and number of poles, an amplifier having a Chebyshev response has considerably more gain than with the maximally flat function, even though the ripples may be small. Actually, the Chebyshev function is a better approximation to the brick-wall amplitude characteristic than is the maximally flat function. However, its phase function is not as linear as that of the maximally flat function, which makes the Chevyshev function relatively unsatisfactory in pulse amplifiers.

If a function contains both p-z, it can be converted from a maximally flat function to an approximate equal-ripple function by multiplying the real parts of both the pole and the zero positions by the same constant, just as with all-pole functions.

Another way to arrive at the all-pole Chebyshev function is to start directly from a Chebyshev polynomial $C_n(x)$, where x is a real variable. Several of these polynomials in normalized form along with the recursion

relation are given by eqs. 5.31

$$\begin{aligned} C_1(x) &= x \\ C_2(x) &= 2x^2 - 1 \\ C_3(x) &= 4x^3 - 3x \\ C_4(x) &= 8x^4 - 8x^2 + 1 \\ C_5(x) &= 16x^5 - 20x^3 + 5x \\ C_{n+1}(x) &= 2xC_n(x) - C_{n-1}(x) \end{aligned} \tag{5.31}$$

When plotted as a function of x, the value of a Chebyshev polynomial oscillates between ± 1 for $|x| < 1$, is equal to ± 1 at $x = \pm 1$, and goes

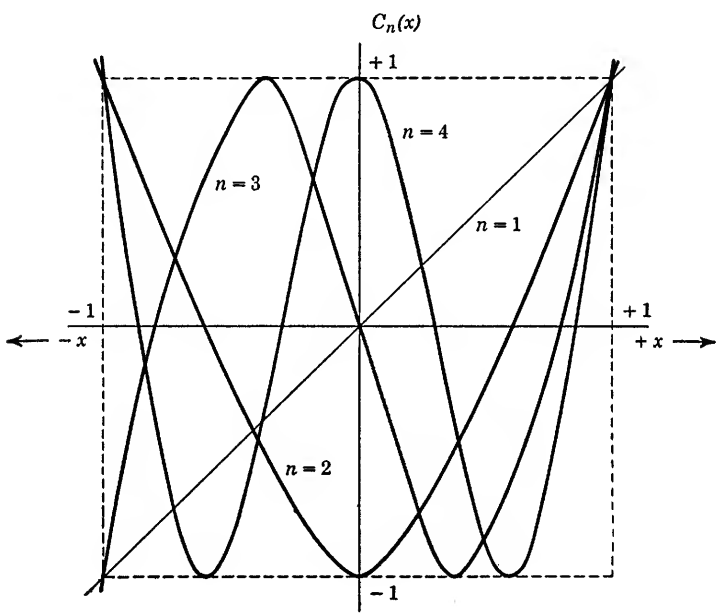


Fig. 5.9. Characteristics of Chebyshev polynomials.

to $\pm\infty$ as x increases beyond ± 1 to $\pm\infty$. The characteristics of the first few polynomials are sketched (not plotted) in Fig. 5.9.

Let us interpret x as the frequency ω and write the square of the magnitude of the transfer function (which has quadrantal symmetry) as

$$F(j\omega)F^*(j\omega) = \frac{1}{1 + \epsilon C_n^2(\omega)} \tag{5.32}$$

Since $C_n^2(\omega)$ is always an even polynomial and is always positive, the value of eq. 5.32 will be equal to or less than unity; it will be equal to unity at values of ω where $C_n(\omega)$ goes to zero and will be equal to $1/(1 + \epsilon)$ where $C_n(\omega) = \pm 1$. In fact, it is not difficult to see that a plot of eq. 5.32 appears as in Fig. 5.10, where the total number of ripples (for both positive and negative frequency ω) is equal to n. Note

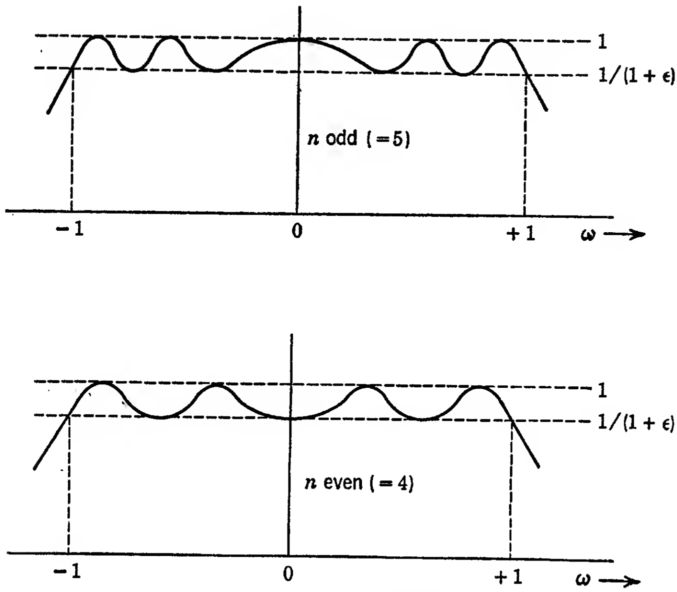


Fig. 5.10. Chebyshev power gain functions.

that for n odd a ripple peak occurs at $\omega = 0$, whereas for n even a ripple trough occurs at $\omega = 0$.

Unlike other functions with which we have dealt, the bandwidth when using the Chebyshev polynomial directly is generally measured from $\omega = 0$ to the limits of the "ripple tolerance" of $\omega = 1$. In other words, within the band $B = 1$, the magnitude of the power gain function is limited between 1 and $1/(1 + \epsilon)$. The half-power bandwidth is slightly larger than unity; it is $\cosh(a)$, where a is given by eq. 5.30.

Setting $\omega = p/j$ in eq. 5.32 converts the equation to a function of p, which has quadrantal symmetry. The poles of the corresponding voltage transfer function for n odd are at

$$p_k = \sinh(-a + jk\pi/n), \quad k = 0, \quad \pm 1, \pm 2, \cdots \pm (n-1)/2 \tag{5.33}$$

and for n even, the poles are at

$$p_k = \sinh[-a + j(1 + 2k)\pi/2n], \quad k = \begin{cases} 0, 1, 2, \cdots (n/2 - 1) \\ -1, -2, \cdots -n/2 \end{cases} \tag{5.34}$$

We have given examples of approximating a brick-wall function in the Taylor sense and again in the Chebyshev sense. These two approximating methods are not limited to brick-wall functions but can be made to approximate any reasonable amplitude function. The Taylor

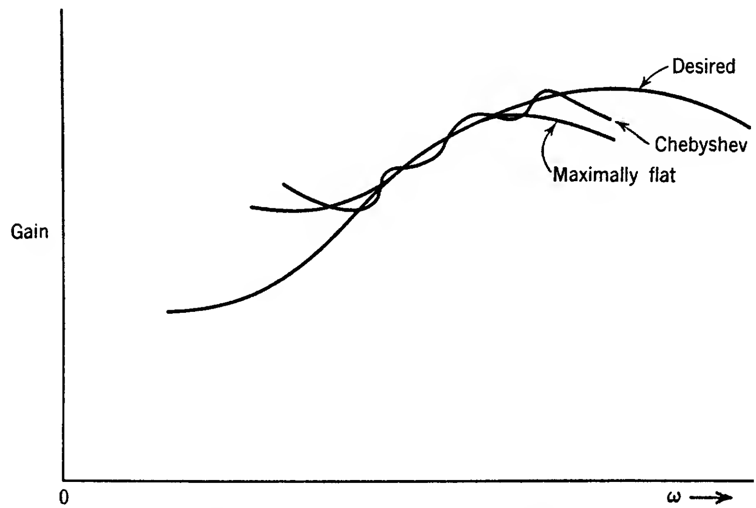


Fig. 5.11. More general approximations.

and Chebyshev approximations of a more arbitrary function might appear somewhat as in Fig. 5.11.

5.5 Linear-phase and compromise approximations

Another approach to realizing the brick-wall function is that of making the function have a linear phase shift with frequency over the band of frequencies containing the principal components of the signal to be transmitted by the function. In the brick-wall gain approximation, the phase shift is ignored. With a linear-phase approximation, the gain function is ignored.

It appears that if a string of poles (an infinite number of them) are equally spaced along a line parallel to the $j\omega$ axis as in Fig. 5.12, the phase shift with frequency will ripple slightly about the desired linear curve. The closer are the poles to the $j\omega$ axis compared to the spacing between poles, the larger will be the magnitude of the ripples.

An approximation to the pole locations of Fig. 5.12 is had by using only a finite part of the picture, as is indicated in Fig. 5.13. This function has some desirable transient properties.

We may combine (in an intuitive manner) the benefits of the maximally flat function and the linear-phase function by placing poles on a

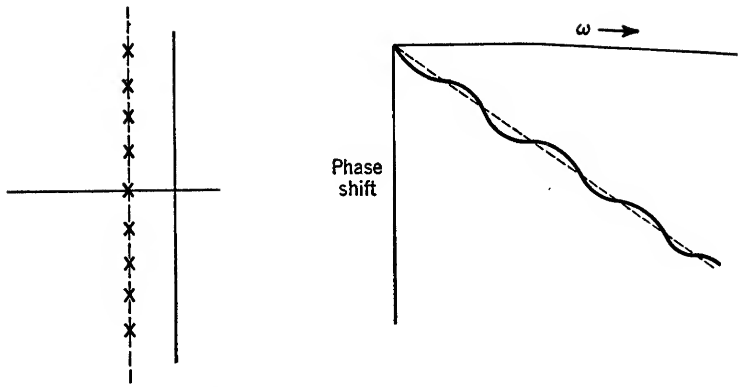


Fig. 5.12. A linear-phase function.

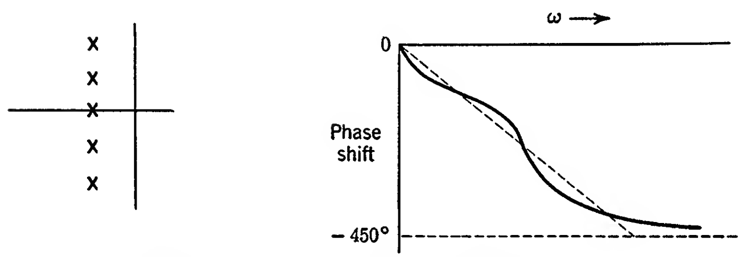


Fig. 5.13. Approximate linear-phase function.

circle about the origin with equal vertical spacings between the poles rather than with equal angles between poles. This is shown in Fig. 5.14.

Another kind of linear-phase approximation is worth deriving here. An amplitude function of the type exp $(-j\omega)$ within the pass band of

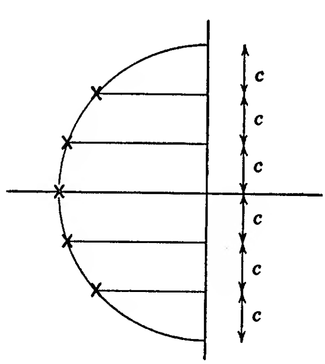


Fig. 5.14. Heuristic approximation of both phase and amplitude.

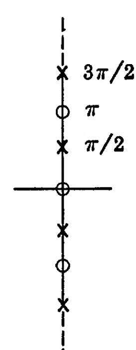


Fig. 5.15. The tangent function of the exponential.

the filter is ideal. The phase shift of this ideal function is $\theta = \tan^{-1}(-\omega)$. The tangent function of the phase angle has an infinite number of p-z on the $j\omega$ axis as shown by Fig. 5.15, in which a pole exists wherever the angle is an odd multiple of $\pi/2$ radians, and a zero wherever the angle is an even multiple of $\pi/2$ radians. The p-z must be separated by $\pi/2$ and must alternate.

The infinite product p equation for tan (θ) is the tangent function described earlier

$$F_p\ (\tan\theta) = \frac{p \prod_{k=1}^{\infty} [1 + p^2/(k^2\pi^2)]}{\prod_{j=1}^{\infty} \{1 + p^2/[(2j-1)\pi/2]^2\}} \tag{5.35}$$

An approximation to this function is one where the infinite limits on k and j are replaced by finite limits. Then at least at the lower frequency points at which the ideal phase function is a multiple of $\pi/2$ radians, there is perfect agreement between the exact and approximate phase expressions.

As an example, consider the two-pole voltage transfer function

$$V = \frac{1}{1 + bp + cp^2} = \frac{1}{1 + bj\omega + c(j\omega)^2} \tag{5.36}$$

which is to be caused to have p-z for the tangent function as in Fig. 5.16. Then

$$\theta = \arg V = \tan^{-1}\left(-\frac{b\omega}{1 - c\omega^2}\right) \tag{5.37}$$

and

$$\tan(-\theta) = \frac{b\omega}{1 - c\omega^2} \tag{5.38}$$

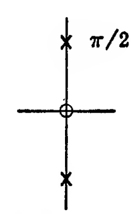


Fig. 5.16. An approximation to the tangent function.

Equation 5.38 can be equated to the corresponding part of eq. 5.35 for $p = j\omega$ to yield

$$\frac{\omega}{1 - \omega^2/(\pi/2)^2} = \frac{b\omega}{1 - c\omega^2} \tag{5.39}$$

By equating coefficients, we get $c = (2/\pi)^2$ and $b = 1$ to give the transfer function as

$$V = \frac{1}{1 + p + (2p/\pi)^2} \tag{5.40}$$

By normalizing eq. 5.40 with respect to frequency so that the coefficient of p^2 is unity, we get

$$V = \frac{1}{1 + \pi p/2 + p^2} \tag{5.41}$$

which has a half-power bandwidth of 0.79 radian. The pole locations are

$$\begin{aligned} p_1, p_1{}^* &= -(\pi/4) \pm j[1 - (\pi/4)^2]^{1/2} \\ &= -0.786 \pm j0.618 \end{aligned} \tag{5.42}$$

These pole positions are compared to the pole positions for the two-pole maximally flat function in Fig. 5.17, from which it can be observed that the magnitude of the linear-phase function decreases more gradually beginning at slightly lower frequencies than does that of the maximally flat function. The approximately equal vertical spacing between the poles is like that of the heuristic function discussed before.

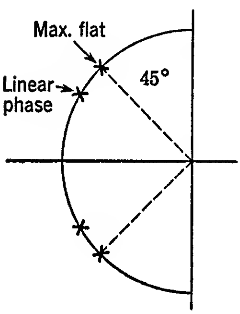


Fig. 5.17. Comparison to the maximally flat function.

The example given here is a very simple one; however, the method is general and can be applied in exactly the same manner for more complex all-pole functions or for functions containing both p-z.

One final linear-phase function of considerable importance makes use of the Taylor approximation to the ideal brick-wall phase function $\theta_1 = -a\omega$. The mathematics is quite similar to that employed for approximating the brick-wall amplitude characteristic. Although the procedure to be outlined is useful for functions having both p-z, we shall restrict our interest to one example, that of a three-pole function. However, the methods to be demonstrated are general.

Ideally, the phase shift should be $\theta_1 = -a\omega$. Then

$$\frac{d\theta_1}{d\omega} = -a \tag{5.43}$$

and all higher derivatives of θ_1 are zero. The unnormalized three-pole function can be described by

$$F(p) = \frac{1}{1 + b_1 p + b_2 p^2 + b_3 p^3} \tag{5.44}$$

from which θ_2, its phase angle, is given by

$$\theta_2 = \tan^{-1}\left(-\frac{b_1\omega - b_3\omega^3}{1 - b_2\omega^2}\right) \tag{5.45}$$

and

$$\frac{d\theta_2}{d\omega} = -b_1\left(\frac{1 + \omega^2(b_1b_2 - 3b_3)/b_1 + \omega^4(b_2b_3/b_1)}{1 + \omega^2(b_1^2 - 2b_1) + \omega^4(b_2^2 - 2b_1b_3) + \omega^6 b_3^2}\right) \tag{5.46}$$

Clearly, in order that eq. 5.46 approximate eq. 5.43, it is first necessary that $b_1 = a$. For a first-order maximally flat phase angle, the coefficients of ω^2 in numerator and denominator of eq. 5.46 must be equal. This gives

$$\frac{b_1b_2 - 3b_3}{b_1} = b_1^2 - 2b_2 \tag{5.47}$$

For a second-order maximally flat phase angle, we must have in addition

$$\frac{b_2b_3}{b_1} = b_2^2 - 2b_1b_3 \tag{5.48}$$

Let us normalize to $a = b_1 = 1$. Then solving for b_2 and b_3, the transfer function becomes

$$F(p) = \frac{1}{1 + p + 0.4p^2 + 0.0667p^3} \tag{5.49}$$

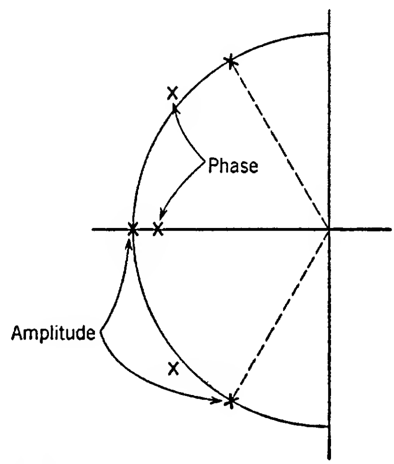


Fig. 5.18. Maximally flat amplitude and phase functions.

Let us now unnormalize so that the coefficients of the highest and lowest powers of p are unity. It should then be easier for the reader to compare the results to the maximally flat three-pole transfer function. We get

$$F(p) = \begin{cases} \dfrac{1}{1 + 2.465p + 2.43p^2 + p^3} & \text{(Max. flat phase)} \\ \dfrac{1}{1 + 2p + 2p^2 + p^3} & \text{(Max. flat amplitude)} \end{cases} \tag{5.50}$$

The poles of eqs. 5.50 are sketched in Fig. 5.18. Again notice the approximately equal vertical spacing of the poles of the maximally flat

phase function. The half-power bandwidth of this maximally flat phase function is, of course, somewhat less than unity (about 0.71).

5.6 Quasi-distortionless functions

The ideal exponential all-pass transfer function has unity gain for all frequencies and a phase angle that increases negatively and linearly with frequency. Expanding the negative exponential in a power series gives

$$e^{-pT} = \frac{1}{e^{pT}} = \frac{1}{1 + pT + (pT)^2/2! + (pT)^3/3! + \cdots} \tag{5.51}$$

If the higher terms of this series $(pT)^n/n! = (j\omega T)^n/n!$ are small compared to the lower terms over the frequency range containing signal components of interest, then an approximation to the ideal exponential valid at the lower frequencies is had by retaining only the first few terms of the expansion. The first-order approximation is

$$\frac{1}{1 + pT} \tag{5.52}$$

which is a simple one-pole low-pass transfer function having a bandwidth of $1/T$. The second-order approximation is

$$\frac{1}{1 + pT + (pT)^2/2} \tag{5.53}$$

which is a two-pole function that also is the maximally flat amplitude function with a half-power bandwidth of $(2)^{1/2}/T$. The third-order approximation is

$$\frac{1}{1 + pT + (pT)^2/2 + (pT)^3/6} \tag{5.54}$$

which is a low-pass function having a slight peak at high frequencies and a half-power bandwidth of $2.16/T$. The magnitude of eq. 5.54 is unity at $\omega = 0$ and also at $\omega = (3)^{1/2}/T$. The fourth-order all-pole approximation has a severe high-frequency peak and does not appear to be too useful. Fifth- and higher order approximations are useless because they have poles in the right half-plane. As will be discussed shortly, by admitting zeros the number of different and useful functions is considerably increased.

When an input signal is subjected to the ideal exponential transfer function exp $(-pT)$, it is simply delayed by a time T without suffering distortion of any kind. Therefore, the approximation to the exponential can be expected to furnish a delay of T seconds to any complex input

signal. As long as all the important components of the signal occur at frequencies well within the bandwidth of the function, the signal will not be appreciably distorted in passing through the network. For this reason, the approximations to the exponential just given are called "quasi-distortionless" functions, or more simply, Q-D functions. The first two approximations we have obtained give us nothing new, although a different criterion for arriving at the two-pole maximally flat function is demonstrated. The third-order approximation is a different and useful function.

If a transfer function $e^{(p\tau)}$ could be realized, the output from the network would lead the input by a time τ seconds; in other words, the network would be able to predict what the input will be some time in the future. However, an approximation to $e^{(p\tau)}$ is a transfer function containing only zeros, which is not permissible as was demonstrated in Chap. 3. Therefore (and as intuition dictates), the ideal predicting network cannot be realized. We can, however, find value in Q-D functions containing both p-z with the number of poles equal to or larger than the number of zeros. To do this, we approximate the function $e^{(p\tau)}$ times $e^{(-pT)}$. If we make T larger than τ, the network will give a delay of $T - \tau$ seconds, whereas if we make τ larger than T, the network will give a "prediction" (a lead) of $\tau - T$ seconds.

Of course, no network can give an output before an input is applied. The mechanism of prediction with Q-D functions is one in which the network measures the value and first few derivatives of a *continuous* type of input signal and combines these measured values in the proper manner to extrapolate the value of the input into the future.

The general approximation to the Q-D function is evidently

$$\frac{e^{\tau p}}{e^{Tp}} = e^{-(T-\tau)p} = 1 - (T - \tau)p + \frac{(T - \tau)^2 p^2}{2!} + \cdots \tag{5.55}$$

Given some function consisting of the ratio of two polynomials in p, we must seek to expand the function in the form of eq. 5.55 and then equate the coefficients of the powers of p. Let us take a general rational transfer function as

$$F(p) = H\,\frac{1 + a_1 p + \cdots + a_m p^m}{1 + b_1 p + \cdots + b_n p^n} = H\,\frac{1 + P_1(p)}{1 + P_2(p)} \tag{5.56}$$

in which we have defined the polynomials P_1 and P_2 and where the gain is finite at zero frequency.

If $P_2(p) = P_2(j\omega)$ is smaller than unity over the range of frequencies occupied by the signal passing through the network, we may employ the expansion

$$\frac{1}{1 \pm x} = 1 \mp x + x^2 \mp x^3 + x^4 \mp \cdots \tag{5.57}$$

to obtain

$$F(p) = H(1 + P_1)(1 - P_2 + P_2^{\,2} - P_2^{\,3} + \cdots) \tag{5.58}$$

Substituting the polynomials for P_2 and P_1 in eq. 5.58 and collecting the coefficients of the powers of p, we get

$$\begin{aligned} F(p) = H[1 &- (b_1 - a_1)p + (a_2 + b_1^{\,2} - a_1b_1 - b_2)p^2 \\ &- (b_3 + b_1^{\,3} - 2b_1b_2 - a_1b_1^{\,2} + a_1b_2 + a_2b_1 - a_3)p^3 \\ &+ \text{higher powers of } p] \end{aligned} \tag{5.59}$$

Immediately, we see by comparing eqs. 5.59 and 5.55 that

$$b_1 - a_1 = T - \tau \tag{5.60}$$

Therefore, a transfer function that is finite at $\omega = 0$ can *always* be written

$$F(p) = \frac{1 + \tau p + a_2p^2 + a_3p^3 + \cdots}{1 + Tp + b_2p^2 + b_3p^3 + \cdots} \tag{5.61}$$

Consider the function having one pole and one zero. The pole and zero of this function arranged to provide lead and delay are shown in Fig. 5.19. These simple functions should be quite familiar by now and have been shown to be related to the Q-D concept.

Fig. 5.19. Simple delay and lead functions.

With more than one pole and one zero, the approximation to the exponential can be improved. For example, consider a function having one zero and two poles. Comparing eqs. 5.55 and 5.59 and setting $a_1 = \tau$, $b_1 = T$, $a_2 = a_3 = \cdots = 0$, and $b_3 = b_4 = \cdots = 0$, we get by

comparing coefficients of p^2

$$b_1^2 - a_1 b_1 - b_2 = (T - \tau)^2/2 \tag{5.62}$$

which can be solved for b_2 to give

$$b_2 = (T^2 - \tau^2)/2 \tag{5.63}$$

The resulting transfer function is

$$F(p) = \frac{1 + \tau p}{1 + Tp + [(T^2 - \tau^2)p^2]/2} \tag{5.64}$$

For $T > \tau$, this function is the maximally flat shunt-peaked function, which is as good an approximation to the exponential as is the two-pole no-zero Q-D function. For $\tau = 0$, it reduces to the two-pole Q-D function (which is also maximally flat). For $\tau > T$, the function cannot be exactly realized. Then, $T^2 - \tau^2$ must more or less arbitrarily be taken positive. As a prediction network function, this function is no better than the simple arrangement of Fig. 5.19*b*. In general, a Q-D prediction function can be made no better an approximation to the exponential than the number of terms in the numerator of the approximating function (the number of zeros) permits. As a delay function with or without zeros, the approximation to the exponential can be made no better than the number of terms in the denominator of the approximating function (the number of poles) permits. In either case, of course, the function must be stable, which means that the poles must all lie in the left half-plane. A delay function with zeros can be a better approximation to the exponential than one without zeros because the function can have more than four poles and still be stable.

If the expansions defining the general Q-D function having m zeros are followed through, it is found that the first m coefficients in both numerator and denominator separately follow the expansion for the exponential. As an example consider eq. 5.65, which is that of a Q-D function having three zeros and five poles

$$F(p) = H \frac{1 + \tau p + (\tau p)^2/2 + (\tau p)^3/6}{1 + Tp + (Tp)^2/2 + (Tp)^3/6 + b_4 p^4 + b_5 p^5} \tag{5.65}$$

For delay networks, only the coefficients b_4 and b_5 need be determined by comparing the series expansions of eqs. 5.55 and 5.59. For lead networks, b_4 and b_5 must be taken more or less arbitrarily to give reasonable pole locations.

Example p-z plots of delay Q-D transfer functions are shown in Fig. 5.20. Realizable Q-D lead functions are shown in Fig. 5.21.

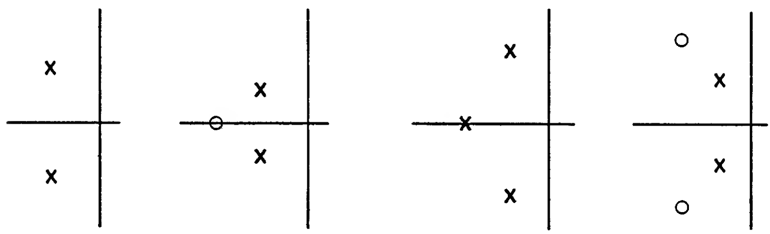


Fig. 5.20. Q-D delay functions.

The ideal differentiating network has a transfer function of simply p. Unfortunately, this function cannot be precisely realized. A practical function is one where the ideal derivative function is followed by a Q-D

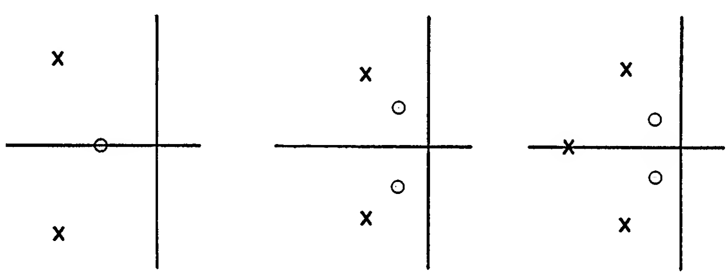
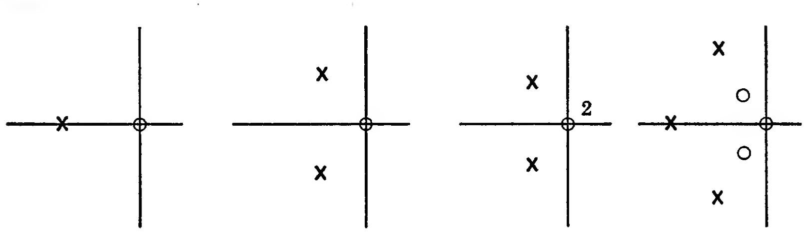


Fig. 5.21. Q-D lead functions.

function to give a network whose output is a delayed or predicted value of the derivative of the input, as exemplified by the Q-D functions shown in Fig. 5.22. (As a general rule, any low-pass function plus a zero at the origin is an ideal derivative network in cascade with a delay network.)



(a) *(b)* *(c)* *(d)*

Fig. 5.22. Derivative functions. (*a*) and (*b*) With delay. (*c*) Second derivative with delay. (*d*) With lead.

5.7 The method of trial and error

Perhaps the most powerful and certainly the most general method of function design is that provided by trial and error. This method is not at all restricted to low-pass functions; its main value is in the design of equalizing functions. Because of the phasor interpretation of the p-plane, many reasonable amplitude or phase functions can be designed by manipulating a few p-z over the p plane. In doing this manipulation, enough p-z should be used to provide the basic requisites of the desired transfer function such as the off-channel rejection, flatness of band, and so forth. Of course, the basic laws of functions must be obeyed: that is, p-z must occur as complex conjugates if complex; no poles can occur on or to the right of the $j\omega$ axis; and there must be the same number or more poles than zeros, usually more. In addition, if nonminimum-phase functions are not desired, no zeros can occur on or to the right of the $j\omega$ axis.

It must also be remembered that, since the amplitude function of a minimum-phase network specifies the phase function, it is not possible to approximate independently both phase and amplitude with minimum-phase functions.

The reader should not minimize the power of the trial and error method. There is a tendency to look down upon it because it does not take the form of a mathematical nicety. Such a viewpoint is not at all justified. The main disadvantage of the method of trial and error is that calculations may become somewhat tedious, although analog computing (as with an electrolytic tank) can ease the tedium considerably.

Problems

1. Explain why an infinite number of poles is needed in order to give the ideal brick-wall function.

2. Prove that the first $2n - 1$ derivatives of $1/(1 + \omega^{2n})$ with respect to ω are zero at $\omega = 0$.

3. Plot $1/(1 + \omega^{2n})^{1/2}$ for $n = 1, 2, 3, 4$, and 5.

4. Obtain the pole locations for a maximally flat three-pole function having a bandwidth B. By equating coefficients, determine the element values of the circuit

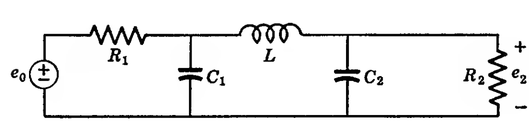


Fig. P.4.

of Fig. P.4 in order that the transfer function be maximally flat and have a bandwidth B; that is, determine C_1, C_2, and L in terms of R_1, R_2, and B. It is helpful to normalize one of the resistors to unity.

5. Referring to Prob. 4, find C_1, C_2, and L when $R_1 = R_2 = 1\text{K}$ and $B = 1$ mcs.

6. Determine L and C for the two-pole network of Fig. P.6 in terms of R_1, R_2, and B in order that the transfer function be maximally flat.

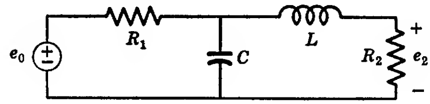


Fig. P.6.

7. Given the function

$$\frac{1 + a_1 p}{1 + b_1 p + b_2 p^2 + b_3 p^3}$$

determine the relations between the coefficients in order that the function be maximally flat. Do not assume any of the coefficients to be zero.

8. Repeat Prob. 7 for the function

$$\frac{1 + a^2 p^2}{1 + b_1 p + b_2 p^2 + b_3 p^3}$$

and plot the locations of its p-z for $a = 0.5$ and $b_3 = 1$. Also, sketch the amplitude and phase as functions of frequency.

9. Determine the element values for the circuit of Fig. P.9 in order that its input impedance be the maximally flat function of Prob. 8. (Note: This circuit cannot be made to realize all values of the coefficients.)

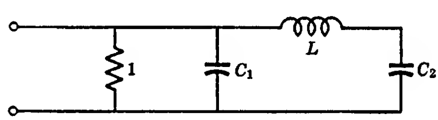


Fig. P.9.

10. Determine C_1, C_2, and L for the circuit of Fig. P.10 in order that its transfer function be the maximally flat function of Prob. 8. (Note: This circuit cannot be made to realize all values of the coefficients.)

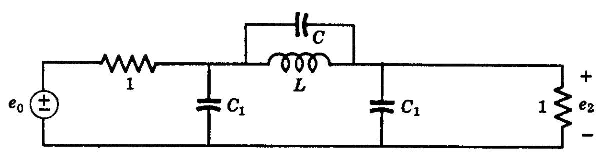


Fig. P.10.

11. Derive the bandwidth-narrowing equation for the shunt-peaked circuit for $\delta = 1$ and also for $\delta = 1.554$.

12. Draw curves for B_N/B (bandwidth narrowing) for all-pole maximally flat functions with one, two, three, four, and five poles and for a few values of N in the region 1 to 10.

13. Determine the pole locations for a maximally flat three-pole function with $B = 1$. Then, determine the pole locations for a three-pole Chebyshev function hav-

ing the same half-power bandwidth and for $\epsilon = 0.1, 0.2$, and 0.3. In each case, form the transfer function as a polynomial and evaluate the coefficients.

14. Evaluate the circuit elements of Fig. P.14 in order that the transfer function be a Chebyshev function with a half-power bandwidth of unity and for $\epsilon = 0.1$. Compare the values of capacitance to those for the maximally flat function having the same bandwidth.

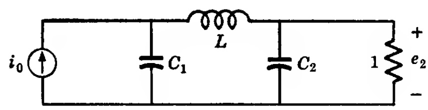


Fig. P.14.

15. Using the results of Prob. 14, obtain elements of the network such that the bandwidth is 8 mcs and has the largest possible terminating resistance. Assume that the minimum permissible values of C_1 and C_2 are 5 and 10 $\mu\mu$f respectively.

16. Making direct use of the Chebyshev polynomial for $n = 3$ and for $\epsilon = 0.05$, determine by equating coefficients the values of C_1, C_2 and L such that the circuit of Fig. P.14 will have a tolerance bandwidth of 1 radian.

17. Determine the pole locations for a three-pole function designed to approximate both the gain and phase characteristics of the brick-wall function. Use a bandwidth of about 1 radian.

18. Determine the coefficients and pole positions of the function

$$\frac{1}{1 + b_1p + b_2p^2 + b_3p^3}$$

in order that its phase approximate that of the exponential. Take the 90-degree phase-shift frequency as unity.

19. Determine the p-z of the shunt-peaked function when the slope of the phase function of frequency is zero at $\omega = 0$. Sketch the magnitude and phase angle of the resulting input impedance and plot the locations of the p-z. ($\delta = 1.554$.)

20. Use the method of trial and error in order to determine the p-z locations for a low-pass function having a bandwidth of unity, frequencies of zero gain at 1.5 and 2, approximately a brick-wall gain function in the pass band, and a gain falling off inversely as the square of frequency at very high frequencies. Use the minimum possible number of p-z. Sketch the gain characteristic of the resulting function.

21. Obtain the coefficients for a Q-D delay function having one zero and three poles.

22. See if you can obtain relationships between bandwidth and delay time for one-, two-, three-, and four-pole maximally flat functions. It will be necessary to use the number of poles n in this "rule of thumb" relationship.

23. Determine element values for the circuit of Fig. P.14 in order that its transfer function be the three-pole Q-D function having a bandwidth of unity. Plot the magnitude of the transfer function against frequency. Compare element values to those obtained when the transfer function is maximally flat and has the same bandwidth.

24. Find the element values for the circuit of Fig. P.24 when the input impedance has maximally flat amplitude, maximally flat phase, exponential linear phase, and is Q-D. Normalize so that the half-power bandwidth is the same for all four cases. Tabulate results and include a comparison of the pole and zero locations.

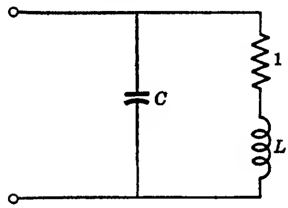


Fig. P.24.

25. A maximally flat function of unity half-power bandwidth has two poles. Determine the step-function response and plot.

26. Move the two poles of the function of Prob. 25 in by multiplying their real parts by 0.9. Determine the amount of ripple, the bandwidth, and the increase in gain. Plot the step-function response with gain normalized to give the same final value as in Prob. 25.

27. Repeat Prob. 26 after moving the maximally flat poles out by multiplying their real parts by 1.2.

28. Compare the step-function response of Probs. 25, 26, and 27. Discuss and generalize. In particular, note the relative amount of overshoot and determine the relationship between rise time and bandwidth.

6

An Introduction to Modern Synthesis

Some topics from advanced network theory are important in the design of practical vacuum-tube and other networks. In particular, a simplified version of the Darlington synthesis and related procedures are of sufficient value to warrant treatment in some detail. This requires a study of general procedures for specifying input immittances.

Any realizable transfer function which has zeros, if any, that reside only on the $j\omega$ axis can be realized with a ladder network containing only inductance and capacitance (and sometimes mutual inductance) which is terminated in a single resistance. The proof that such a network exists (which will not be given here) is one of the most important results of modern theory.

Modern network synthesis is an exact process in that it attempts a precise realization of a network function. This can be contrasted to "classical" (or "conventional") filter design (to be treated in the following two chapters) which obtains filter circuits having characteristics subject to certain idealizations, the resulting physical filter often having a behavior which is only an approximation to that given when the idealizations hold.

6.1 Synthesis of more general input immittances

Several different procedures exist for finding a network that obtains a physically realizable input immittance. We have already treated networks which consist of only two kinds of elements, that is, R-C, R-L, and L-C. It is now our purpose to consider the realization of input immittances to networks containing R, L, and C.

Of all theorems regarding input immittances, the most important from a practical standpoint comes from S. Darlington. The theorem states that the input immittance of any physical bilateral network can always be realized as the input immittance of a four-terminal network containing only L, C, and M (dissipationless) which is terminated in a single resistance. This is a most remarkable theorem which we shall not

attempt to prove here. In pictorial form, it means that any positive-real input immittance can be realized with the circuit of Fig. 6.1.

A general Darlington network is a ladder which may contain four different types of "sections," two of which require mutual inductance. We shall not consider the more complicated structures but will restrict

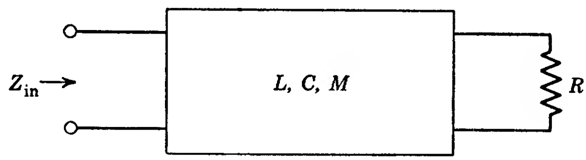


Fig. 6.1. The dissipationless four-terminal network.

ourselves to those containing only L and C. This means that transformers will be included only if they can be replaced with their T or pi equivalents containing positive inductances. With this restriction, most Darlington networks have the form of Fig. 6.2, where each box represents an L-C network. The first and/or last boxes may be missing so

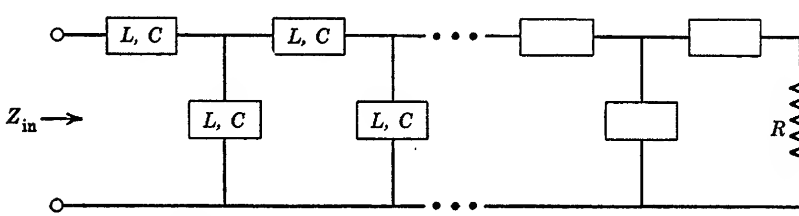


Fig. 6.2. The dissipationless L-C ladder network.

that the beginning and/or ending elements are in shunt rather than in series. Although each box may represent a very complicated L-C network, some of the most practical networks contain individual boxes having one of the forms indicated by the circuits of Fig. 6.3. If a problem cannot be solved with the boxes of Fig. 6.3 or series and parallel

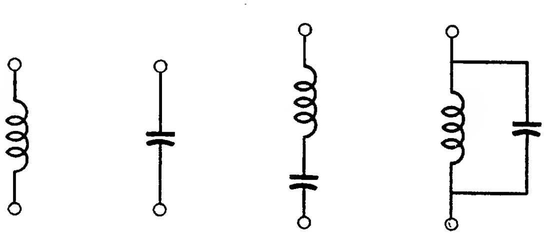


Fig. 6.3. Types of elements in the L-C ladder network.

combinations of such simple networks, it is necessary to use the more complex Darlington sections and hence refer to sources other than this.

By far the most important network is one like that of Fig. 6.4 (or the band-pass, high-pass, or band-elimination equivalent), or in the same category by virtue of the fact that each shunt and series box contains

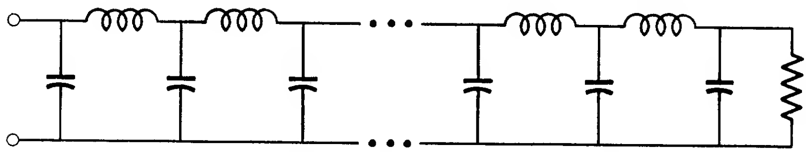


Fig. 6.4. Most used *L-C* low-pass ladder network.

only a single L or C. As an example of obtaining such networks, consider eq. 6.1, which can be assumed to represent a given input admittance

$$Y = \frac{p^3 + p^2 + 2p + 1}{p^2 + p + 1} \tag{6.1}$$

Let us try a continued-fraction expansion starting with the highest power of p. We would not start with the lowest power in this case because that would indicate a resistance at the beginning rather than at the end of the network which, we shall assume, is not desirable for this particular synthesis task. We get

$$Y = p + \cfrac{1}{p + \cfrac{1}{p + 1}} \tag{6.2}$$

which is meaningful and is given by the network of Fig. 6.5.

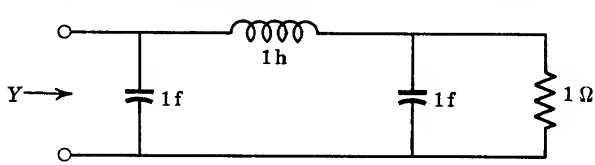


Fig. 6.5. Realization of a simple ladder.

The simple continued-fraction expansion does not always work, particularly when certain combinations of L and C occur in adjacent shunt and series arms. Numerous trials may have to be made in order to determine whether division of the remainder must be done starting with high or low powers of p. The important thing is for the expansion to be meaningful. If the expansion cannot be made to terminate or if nega-

tive terms are given, the simple method described here cannot be used. The example of eqs. 6.3, which is a ladder but which is not a Darlington-type network, indicates some common trial and error procedures

$$
\begin{aligned}
Z &= \frac{2p^3 + 3p^2 + 2p + 1}{2p^2 + 2p + 1} = p + \cfrac{1}{\cfrac{2p^2 + 2p + 1}{p^2 + p + 1}} \\
&= p + \cfrac{1}{\cfrac{1 + 2p + 2p^2}{1 + p + p^2}} = p + \cfrac{1}{1 + \cfrac{1}{\cfrac{p^2 + p + 1}{p^2 + p}}} \\
&= p + \cfrac{1}{1 + \cfrac{1}{\cfrac{1 + p + p^2}{p + p^2}}} = p + \cfrac{1}{1 + \cfrac{1}{\cfrac{1}{p} + \cfrac{1}{\cfrac{p + 1}{p}}}}, \\
&= p + \cfrac{1}{1 + \cfrac{1}{\cfrac{1}{p} + \cfrac{1}{\cfrac{1}{p} + 1}}}
\end{aligned}
\tag{6.3}
$$

It is easy to see that the impedance of eqs. 6.3 can be realized with the circuit of Fig. 6.6.

It is possible to test the validity of a continued-fraction expansion as the design progresses by studying the remainder after each division. If

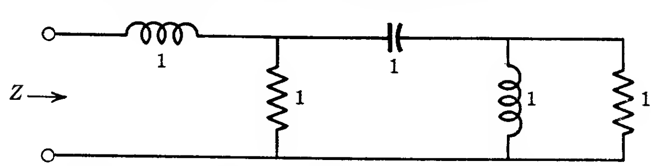


Fig. 6.6. Realization of a simple *R-L-C* impedance.

the process is to give a physical ladder network, each remainder must represent a realizable input immittance; that is, the remainder must be positive real.

Perhaps the most important class of networks, always capable of having its input impedance expressed as a simple continued-fraction expansion, is the ladder network with series L arms and shunt C arms. The transfer function of such networks, when driven with either an ideal or practical voltage or current generator (with resistive source impedance, if any) and when terminated in a finite or infinite resistance, will

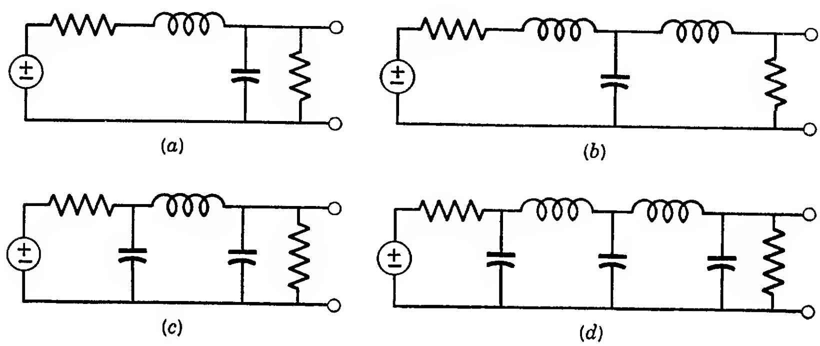


Fig. 6.7. All-pole L-C ladder networks. (*a*) Two poles. (*b*) and (*c*) Three poles. (*d*) Five poles.

always show N poles and no zeros, where N is the total number of inductors and capacitors in the network. The element closest to the generator and/or the load may be either a series L or a shunt C. Examples of such networks are shown in Fig. 6.7. (Note that the networks of Figs. 6.7*b* and *c* are duals.) With their use, any one of a variety of all-pole transfer functions may be realized.

6.2 A more general synthesis of L-C ladder networks

The general ladder network not containing mutual inductance contains series and shunt ladder arms that are two-terminal L-C networks. Thus, each arm is an immittance function having p-z only on the $j\omega$ axis which obey the separation property, and so forth. If the arms of the ladder are fairly complicated, any one of several different L-C networks may be used to represent each ladder arm, that is, one of the four canonical forms or some combination of these forms. For the purposes of the present analysis, let us represent each shunt admittance of the ladder with the canonical form of Fig. 6.8*a* and each series impedance with the form of Fig. 6.8*b*. After the synthesis procedure is complete and the network obtained, different canonical forms may be obtained for the various series and shunt arms of the ladder if desired. The method we shall present here does not always work, even when the net-

work can be found as an *L-C* ladder network. However, it does work in many practical cases.

The synthesis procedure is started by expanding the input admittance of the network of Fig. 6.9 into partial fractions. Several of the terms of this expansion may represent the admittance Y_1. If $Y_1 = 0$, the partial-fraction expansion for the input impedance rather than the admittance should be made and the expansion would then have terms representing the first series arm of the ladder. In order to describe the procedure, it will be assumed that Y_1 is not zero.

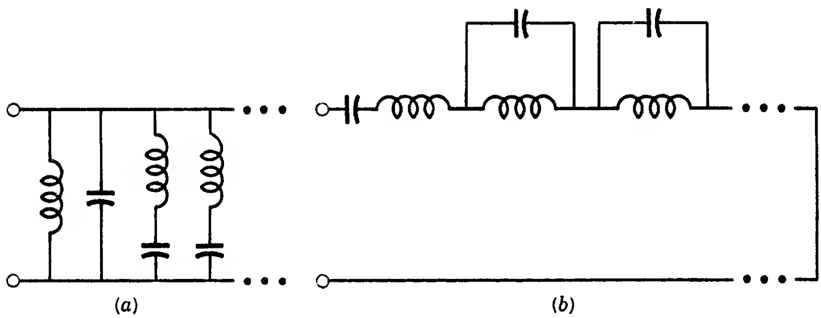


Fig. 6.8. Assumed structure of the ladder arms.

The poles and zeros of Y_1 all lie on the $j\omega$ axis and are simple. The poles (but not the zeros) of Y_{in} may contain those of Y_1. Hence, the terms of the partial-fraction expansion representing Y_1 may be those of Y_{in} that show poles on the $j\omega$ axis. Thus, the terms of the partial-fraction expansion relating to poles on the $j\omega$ axis can be chosen for Y_1 such that the sum of the rest of the terms of the expansion for Y_{in} represents a realizable input admittance.

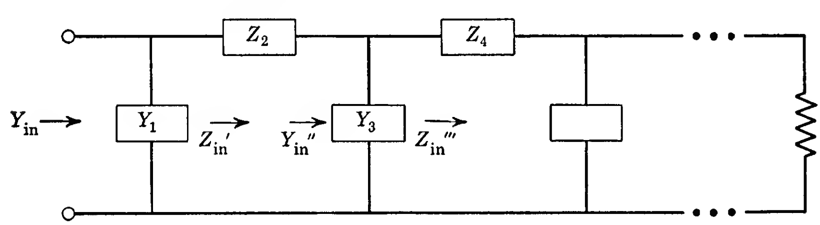


Fig. 6.9. A more general *L-C* ladder network.

This completes the first step of the synthesis in that the first shunt arm Y_1 has been determined. Next, the remaining term of Y_{in}, $Y_{in} - Y_1$, is changed to a rational fraction which is then inverted to obtain the

impedance Z_{in}'. This impedance is expanded into partial fractions. The terms of Z_{in}' showing poles on the $j\omega$ axis may belong to Z_2. Hence, the terms representing Z_2 can be extracted from the expansion in order to determine the structure of Z_2. Then, the inverse of the new remainder $Y_{in}'' = 1/(Z_{in}' - Z_2)$ is expanded into partial fractions and the process is repeated. This step by step evaluation of the arms of the ladder is continued until it (hopefully) terminates in a resistor at the end of the network.

As a very simple example of the procedure just outlined, consider the admittance function of eq. 6.4. This

$$Y_{in} = \frac{2p^2 + p + 1}{(p^2 + 1)(p + 1)} \tag{6.4}$$

function has poles on the $j\omega$ axis but no zeros on the $j\omega$ axis; hence, the leading arm of the ladder is a shunt arm and we expand Y_{in} as

$$Y_{in} = \frac{\frac{1}{2}}{p + j} + \frac{\frac{1}{2}}{p - j} + \frac{1}{p + 1} = \frac{p}{p^2 + 1} + \frac{1}{p + 1} \tag{6.5}$$

Had there been zeros of Y_{in} on the $j\omega$ axis rather than poles, the input impedance rather than the admittance would have poles on the $j\omega$ axis and Z rather than Y would be expanded.

The term $p/(p^2 + 1)$ represents the admittance of a series-resonant circuit. Hence, the network can be drawn as in Fig. 6.10a. The re-

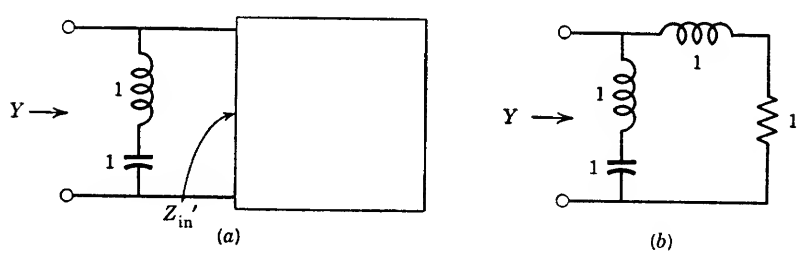


Fig. 6.10. Realization of an L-C network.

mainder is $1/(p + 1)$ which is Y_{in}'. The impedance Z_{in}' is $p + 1$, which is that of a series coil and resistor. The complete network is shown in Fig. 6.10b.

When an immittance expression has no poles or zeros on the $j\omega$ axis at finite frequencies, a modified and simpler procedure is sometimes applicable. If the expression for the input impedance has a numerator that is of higher degree in p than the denominator, the partial-fraction

expansion will have a term Lp, which represents a series L. Similarly, if the numerator of an expression for admittance is of higher degree in p than the denominator, the partial-fraction expansion has a term Cp, which represents a shunt C. As far as this one term of the expansion is concerned, the partial-fraction expansion is the same as the continued-fraction expansion and the rather tedious task of finding the complete partial-fraction expansion in order to find the value of this particular L or C can be avoided.

When a term Lp or Cp appears in a partial-fraction expansion, the function goes to infinity at infinite frequencies and it can be said that the function has a pole at infinity, assuming that infinity lies on the $j\omega$ axis. In an analogous manner, a function having a zero at infinity is one whose value is zero at infinite frequencies, which occurs when the denominator of the function is of higher degree in p than the numerator. The reciprocal of a function having a zero at infinity is a function having a pole at infinity. If the notion of poles or zeros at infinity is adopted, it is evident that all rational functions have equal numbers of p-z.

It is easy to see by inspection when an immittance has a pole at the origin. If an impedance has a pole at the origin, there must be a capacitance in the series arm of the latter. If an admittance has a pole at the origin, there must exist an inductance in the shunt arm. The value of L or C may be obtained by writing the partial-fraction expansion only for the term $1/p$ and the remainder, the complete expansion not being required. If an impedance has a zero at the origin, the admittance has a pole at the origin, which infers that the function should be inverted.

It is generally easiest to consider the complex-conjugate poles or zeros of an input immittance function before the poles or zeros at infinite or zero frequency.

As an example of the procedure which involves most of the considerations so far mentioned, consider the impedance function of eq. 6.6

$$Z(p) = \frac{p^4 + 2p^3 + 3p^2 + p + 1}{p^3 + p^2 + 2p} \tag{6.6}$$

This function has no complex poles or zeros on the $j\omega$ axis. However, it has poles at both $p = 0$ and $p = j\infty$. Thus, there exists either a C or an L as the first series element and a choice exists allowing the designer to make trials and pick the trial leaving a remainder that is realizable. Often, two or more choices may lead to realizable networks. Then, different network realizations of the same impedance function may be obtained.

Let us choose to use the idea of the pole at infinity. We get by divi-

sion starting with the highest powers of p

$$Z(p) = p + \frac{p^3 + p^2 + p + 1}{p^3 + p^2 + 2p} = p + \frac{(p+1)(p^2+1)}{p(p^2+p+2)} \tag{6.7}$$

The second term of eq. 6.7 has a pole at the origin. However, it also has complex zeros at $p = \pm j$, which take precedence. Complex zeros of impedance are complex poles of admittance; thus, the function should be inverted to give an admittance. Doing this and forming the partial-fraction expansion, we obtain

$$Z(p) = p + \cfrac{1}{\cfrac{p(p^2+p+2)}{(p+1)(p^2+1)}} = p + \cfrac{1}{\cfrac{p}{p^2+1} + \cfrac{p}{p+1}} \tag{6.8}$$

The term $p/(p^2 + 1)$ is the admittance of a series L-C circuit in shunt, which determines this arm of the network.

The admittance $p/(p + 1)$ has no complex p-z and no pole at infinity. It does have a zero at the origin; hence, its inverse $(p + 1)/p$ is an impedance having a pole at the origin and thus there must be a capacitance in the next series arm. Finally

$$Z(p) = p + \cfrac{1}{\cfrac{p}{p^2+1} + \cfrac{1}{\cfrac{1}{p} + 1}} \tag{6.9}$$

which is given by the network of Fig. 6.11.

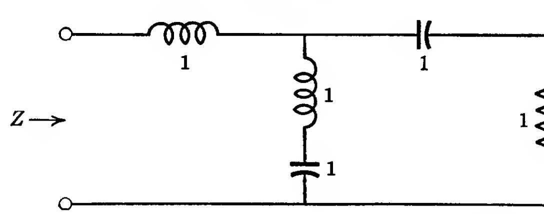


Fig. 6.11. An additional input impedance realization.

In obtaining an L-C ladder network, the poles at $p = 0$ and at infinity and the complex poles on the $j\omega$ axis furnish the key, with complex poles taking precedence for greatest simplicity. After looking for poles on the $j\omega$ axis, poles at infinity should be considered. Zeros at the origin, infinity, or on the $j\omega$ axis indicate that the function should be inverted so that the zeros become poles.

The method for obtaining networks as described in the foregoing is not general and in many instances will not give any realizable circuit even though some L-C ladder network does exist. In this regard, the reader should study the network of Fig. 6.12. If the expression for the input immittance is written (with numerical values) and an attempt is made to synthesize from these data as has been described up to now, failure will result. The synthesis procedures that do exist for the solution of such networks are too involved to consider in this very brief

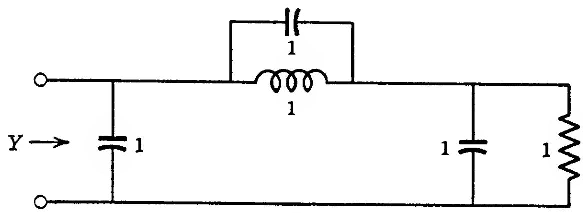


Fig. 6.12. A network yielding some new synthesis problems.

presentation, although we shall say something about the methods as applied to relatively simple cases.

Suppose an input impedance is given for which a Darlington-type network is to be found, for example

$$Y(p) = \frac{3p^3 + 2p^2 + 2p + 1}{2p^2 + p + 1} \tag{6.10}$$

This function has no complex poles or zeros on the $j\omega$ axis or at zero. However, it does have a pole at infinity; that is, the numerator is of higher degree than the denominator. Dividing the denominator into the numerator, we get

$$Y(p) = \frac{3}{2}p + \frac{(p^2/2) + (p/2) + 1}{2p^2 + p + 1} \tag{6.11}$$

which has a remainder that does not have poles or zeros anywhere on the $j\omega$ axis, including zero and infinity. It looks as if we are stuck. Therefore, let us try something different.

The term $3p/2$ represents a $\frac{3}{2}$ f capacitor in shunt at the input to the network. Perhaps it would be better if this shunt capacitor were not quite so large. Therefore, let us divide such that there is obtained something less than $3p/2$, in particular, xp where x is to be determined.

Therefore, we try division as

$$\begin{array}{r|l} & xp \\ \hline 2p^2 + p + 1 & 3p^3 + 2p^2 + 2p + 1 \\ & 2xp^3 + xp^2 + xp \\ \cline{2-2} & (3 - 2x)p^3 + (2 - x)p^2 + (2 - x)p + 1 \end{array} \tag{6.12}$$

which gives

$$Y(p) = xp + \frac{(3 - 2x)p^3 + (2 - x)p^2 + (2 - x)p + 1}{2p^2 + p + 1} \tag{6.13}$$

If we are able to do anything with the remainder, it is necessary that poles or zeros appear somewhere on the $j\omega$ axis. Since the idea of the pole at infinity does not work here and since nothing can be done with the denominator of the remainder, let us choose an x such that the numerator factors to $(ap + 1)(bp^2 + 1)$ so that a pair of zeros appears on the $j\omega$ axis. Comparing the coefficients of the assumed numerator polynomial $abp^3 + bp^2 + ap + 1$ and the numerator of the remainder of eq. 6.13, we find $x = 1$. Therefore

$$Y(p) = p + \frac{(p^2 + 1)(p + 1)}{2p^2 + p + 1} \tag{6.14}$$

From here on, the road is more familiar. We invert the remainder of eq. 6.14 and expand into partial fractions to get

$$Y(p) = p + \frac{1}{\dfrac{2p^2 + p + 1}{(p^2 + 1)(p + 1)}} = p + \frac{1}{\dfrac{p}{p^2 + 1} + \dfrac{1}{p + 1}} \tag{6.15}$$

which is the input admittance of the circuit of Fig. 6.12.

6.3 A special *L-C* ladder synthesis

For regular *L-C* ladder networks of the types shown in Fig. 6.13, it is possible to obtain compact equations for the element values in terms of the coefficients of the powers of p of the input impedance or admittance expression. Let it be assumed that

$$Y_1 \text{ (or } Z_1) = \frac{A}{B} = \frac{a_n p^n + a_{n-1}p^{n-1} + \cdots + a_1 p + a_0}{b_{n-1}p^{n-1} + b_{n-2}p^{n-2} + \cdots + b_1 p + b_0} \tag{6.16}$$

is the input immittance expression for the networks of Fig. 6.13. It is further assumed that all the coefficients a_j and b_j are known. It will be necessary that the immittance expression truly represent the regular

L-C ladder network terminated in a resistor if what we are about to say is to be valid. The element nearest the input end of the ladder may be either a shunt capacitor or a series inductor. Similarly, the element nearest the terminating resistor R_2 may be either a shunt capacitor or a series inductor. The number n, which may be either odd or even, is the total number of inductors and capacitors in the network. Clearly, G_2 (or R_2) $= a_0/b_0$ is the value Y_1 (or Z_1) at $\omega = 0$.

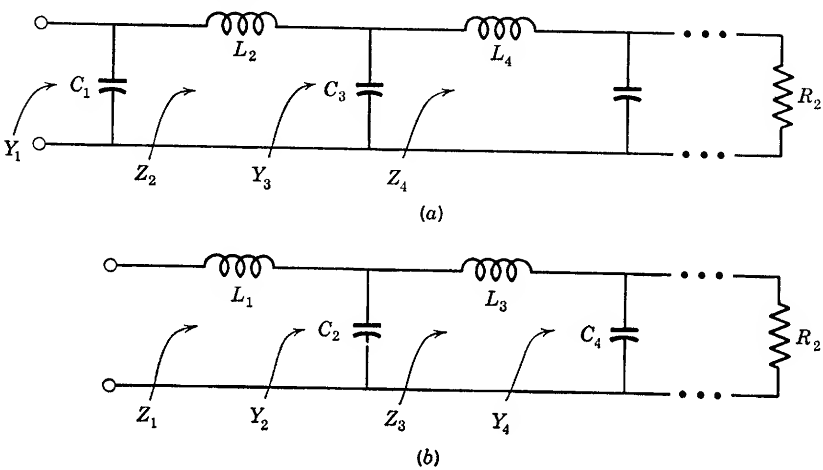


Fig. 6.13. Regular L-C ladder networks.

The first term of the continued-fraction expansion gives the capacitance C_1 (or the inductance L_1) as

$$C_1 \text{ (or } L_1) = \frac{a_n}{b_{n-1}} \tag{6.17}$$

which is a value that can be determined by inspection. Once C_1 has been found, we can form

$$Z_2 = \frac{1}{Y_1 - pC_1} = \frac{B}{A - pC_1B} \tag{6.18}$$

(For simplicity, we shall assume that the leading element is a shunt capacitor from here on. If the circuit actually has a series inductor as a leading element, the method of analysis is essentially the same as that developed here.) The expression for Z_2 has a numerator that is of highest degree $n - 1$ in p. If the leading term of the continued-fraction expansion of Z_2 is to give the element L_2, then the highest degree of p in the denominator must be $n - 2$. Thus, the coefficient of p^n, $a_n - C_1b_{n-1}$,

and of p^{n-1}, $a_{n-1} - C_1 b_{n-2}$, must both be zero. This gives an alternate equation for C_1 as

$$C_1 = \frac{a_{n-1}}{b_{n-2}} \tag{6.19}$$

The value of L_2, as determined from the leading term of the continued-fraction expansion of Z_2, is therefore

$$L_2 = \frac{b_{n-1}}{a_{n-2} - C_1 b_{n-3}} \tag{6.20}$$

After determining C_1 and L_2, we may form

$$\begin{aligned} Y_3 &= \frac{1}{Z_2 - pL_2} = \frac{1}{[B/(A - pC_1B)] - pL_2} \\ &= \frac{A - pC_1B}{B - pL_2(A - pC_1B)} \end{aligned} \tag{6.21}$$

The numerator of Y_3 is the same as the denominator of Z_2. Therefore, the highest power of p in the numerator that does not have a zero coefficient is p^{n-2}. If the leading term of the continued-fraction expansion of Y_3 is to give the capacitance C_3, then the coefficients of all powers of p in the denominator of eq. 6.21 above p^{n-3} must be zero. From this requirement, we get an alternate equation for L_2 as

$$L_2 = \frac{b_{n-2}}{a_{n-3} - C_1 b_{n-4}} \tag{6.22}$$

and the continued-fraction expansion gives

$$C_3 = \frac{a_{n-2} - C_1 b_{n-3}}{b_{n-3} - L_2(a_{n-4} - C_1 b_{n-5})} \tag{6.23}$$

If we continue in this manner, we find regularity clear enough so that the general expression for C can be written

$$C_k = \frac{\text{Denominator of expression for } L_{k-1}}{\begin{pmatrix}\text{Denominator of } C_{k-2} \\ \text{with subscript num-} \\ \text{bers decreased by 2}\end{pmatrix} - L_{k-1} \begin{pmatrix}\text{Denominator of } L_{k-1} \\ \text{with subscript num-} \\ \text{bers decreased by 2}\end{pmatrix}} \tag{6.24}$$

The expression for L_k is the same as eq. 6.24 if all the L and C symbols are interchanged. The alternate expressions for the element values are given by eq. 6.24 if all subscript numbers on a_j and b_j (not on L and C)

are decreased by one, for example, $a_{k-1} \to a_{k-2}$. (An alternate solution does not apply to the element adjacent to the load R_2.)

For any practical network having a finite value of n, the equations for the element values simplify considerably because $a_{n-k} = b_{n-k} = 0$ for $k > n$.

As a specific example, let us find the elements of a three-pole network terminated in R_2 and with shunt capacitors C_1 and C_3 adjacent to both the input and load ends. Then

$$Y_1 = \frac{a_3p^3 + a_2p^2 + a_1p + G_2}{b_2p^2 + b_1p + 1} \tag{6.25}$$

We find

$$\begin{aligned} C_1 &= \frac{a_3}{b_2} = \frac{a_2}{b_1} \\ L_2 &= \frac{b_2}{a_1 - C_1} = \frac{b_1}{G_2} \\ C_3 &= a_1 - C_1 \end{aligned} \tag{6.26}$$

where the alternate solutions show specifically what the relationships between the coefficients a_j and b_j must be in order that the admittance expression correspond to the regular ladder structure.

6.4 Real and imaginary parts of input immittances

Suppose the real part of a bilateral input immittance is given as data and it is desired to find the complete expression for the immittance. A method propounded by Gewertz not only permits this to be done but also proves that the imaginary part is specified uniquely in terms of the real part and conversely, except for an arbitrary reactance or susceptance as will be discussed later.

Let an input impedance be given as the ratio of two polynomials in p as

$$Z(p) = \frac{A(p)}{B(p)} = \frac{a_0 + a_1p + a_2p^2 + \cdots}{b_0 + b_1p + b_2p^2 + \cdots} \tag{6.27}$$

We can define even and odd parts according to

$$\begin{aligned} \text{Ev } B(p) &= b_0 + b_2p^2 + b_4p^4 + \cdots \\ \text{Od } B(p) &= b_1p + b_3p^3 + b_5p^5 + \cdots \end{aligned} \tag{6.28}$$

It should be observed that for $p = j\omega$, $\text{Ev } F(p) = \text{Re } F(j\omega)$ and $\text{Od } F(p) = j \text{ Im } F(j\omega)$, where $F(p)$ is any arbitrary polynomial in p.

Also when $p = j\omega$, $|B(p)|^2 = B(p)B^*(p)$, Ev $B(p) =$ Ev $B^*(p)$, and Od $B^*(p) = -$Od $B(p)$. Thus, we can develop $Z(p)$ as follows, where it is assumed that evaluations are made for $p = j\omega$

$$Z(p) = \frac{A(p)}{B(p)} = \frac{A(p)B^*(p)}{B(p)B^*(p)} = \frac{(\mathrm{Ev}\,A + \mathrm{Od}\,A)(\mathrm{Ev}\,B^* + \mathrm{Od}\,B^*)}{BB^*}$$

$$= \frac{(\mathrm{Ev}\,A + \mathrm{Od}\,A)(\mathrm{Ev}\,B - \mathrm{Od}\,B)}{BB^*} \tag{6.29}$$

$$= \frac{\mathrm{Ev}\,A\ \mathrm{Ev}\,B - \mathrm{Od}\,A\ \mathrm{Od}\,B}{BB^*} + \frac{\mathrm{Od}\,A\ \mathrm{Ev}\,B - \mathrm{Ev}\,A\ \mathrm{Od}\,B}{BB^*}$$

The first term is the real part of $Z(p)$ at $p = j\omega$ and the second term is j times the imaginary part. Thus when the real part of $Z(p)$ is given, we can say

$$\mathrm{Ev}\,Z(p) = \frac{\mathrm{Ev}\,A\ \mathrm{Ev}\,B - \mathrm{Od}\,A\ \mathrm{Od}\,B}{BB^*} \tag{6.30}$$

Equation 6.30 is one showing quadrantal symmetry; hence, the poles of Ev $Z(p)$ are those of $Z(p)$ plus their negatives. Consequently, we select only those poles of Ev $Z(p)$ that appear in the left half-plane as those belonging to $Z(p)$, in which case the determination of the poles of $Z(p)$ knowing only the real part of $Z(p)$ is complete.

The given data are assumed to be

$$\mathrm{Ev}\,Z(p) = \frac{c_0 + c_2p^2 + c_4p^4 + \cdots}{d_0 + d_2p^2 + d_4p^4 + \cdots} \tag{6.31}$$

which must be equal to eq. 6.30. In particular, the numerator of eq. 6.30, with substitution of values from eq. 6.27, must be equal to the numerator of eq. 6.31. Thus

$$c_0 + c_2p^2 + \cdots = (a_0 + a_2p^2 + \cdots)(b_0 + b_2p^2 + \cdots) - (a_1p + a_3p^3 + \cdots)(b_1p + b_3p^3 + \cdots) \tag{6.32}$$

In this equation, the c's are given data. The b's are determined from the locations of the poles. This leaves only the a's as unknowns, which can be evaluated by equating the coefficients of the powers of p term by term.

Much of the preceding can be clarified by means of an example. As given data, assume that

$$\operatorname{Re} Z(j\omega) = 1/(1 + \omega^4) \tag{6.33}$$

or

$$\operatorname{Ev} Z(p) = \frac{1}{1 + p^4} = \frac{1}{(p - p_1)(p - p_1^*)(p - p_2)(p - p_2^*)} \tag{6.34}$$

where p_2 is the negative of p_1 and p_1^* is the conjugate of p_1. The poles are those of a maximally flat function and hence are located on the unit circle with equal angular spacing. The poles of eq. 6.33 are at

$$\begin{aligned} p_1, p_1^* &= -1/(2)^{1/2} \pm j/(2)^{1/2} \\ p_2, p_2^* &= +1/(2)^{1/2} \pm j/(2)^{1/2} \end{aligned} \tag{6.35}$$

It is obvious which poles belong to $Z(p)$. Since $Z(p)$ has only two poles, there can be three zeros at most. Thus

$$Z(p) = \frac{A(p)}{(p - p_1)(p - p_1^*)} = \frac{a_0 + a_1 p + a_2 p^2 + a_3 p^3}{p^2 + (2)^{1/2} p + 1} \tag{6.36}$$

The numerator of $Z(p)$ is found by equating powers of p in eq. 6.32. For that of p^0, $c_0 = a_0 b_0$. But $c_0 = 1$ from the given data and $b_0 = 1$ from eq. 6.36; thus $a_0 = 1$.

Equating coefficients of p^2

$$c_2 = a_0 b_2 + b_0 a_2 - a_1 b_1 \tag{6.37}$$

From the data, $c_2 = 0$. From eq. 6.36, $b_0 = b_2 = 1$ and $b_1 = (2)^{1/2}$. a_0 has already been determined as unity. Thus, eq. 6.37 becomes

$$0 = 1 + a_2 - a_1 (2)^{1/2} \tag{6.38}$$

Equating coefficients of p^4

$$c_4 = a_2 b_2 + a_0 b_4 + b_0 a_4 - b_1 a_3 - b_3 a_1 \tag{6.39}$$

From the given data, $c_4 = 0$. From the poles $b_3 = b_4 = 0$. For realizability, $a_4 = 0$. Thus, eq. 6.39 becomes

$$0 = a_2 - a_3 (2)^{1/2} \tag{6.40}$$

Equating coefficients of p^6 and substituting values gives the relation $a_3 b_3 = 0$. This does not tell us much because we know $b_3 = 0$. Similarly, equating the rest of the coefficients will not tell us anything. Apparently, we have some choice in the matter in that a_1, a_2, and a_3 are not uniquely specified. Let us assume that $Z(p)$ cannot become infinite at infinite frequencies. (The reason for this assumption will be ex-

plained shortly.) Then, the number of zeros of $Z(p)$ cannot be greater than the number of poles, which requires that $a_3 = 0$ and permits us to determine a_1 and a_2 uniquely to give

$$Z(p) = (\tfrac{1}{2})^{1/2}\left(\frac{p + (2)^{1/2}}{p^2 + (2)^{1/2}p + 1}\right) \tag{6.41}$$

which is the input impedance of the shunt-peaked circuit of Fig. 6.14 having values of R, L, and C causing it to be maximally flat with unity bandwidth.

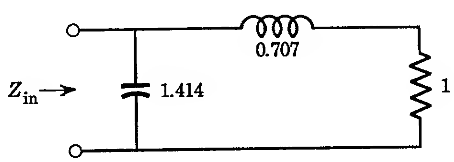


Fig. 6.14. Network obtained knowing the real part of Z_{in}.

For the foregoing example, a completely defined impedance function could not be found from a knowledge of the real part without specifying that the number of zeros could not exceed the number of poles. In order to explain the reason for this, consider Fig. 6.15, which shows a

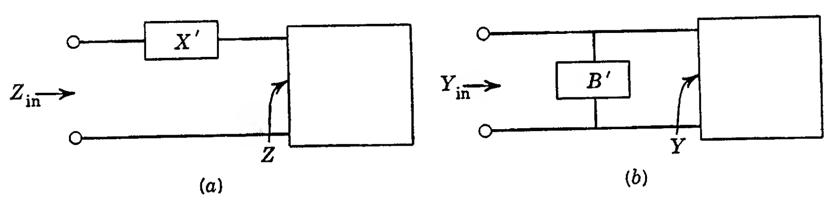


Fig. 6.15. Excess immittances.

reactance X' or susceptance B' external to the general immittance Z or Y. Let $Z = R + jX$ and $Y = G + jB$. Then

$$\begin{aligned} Z_{in} &= R + j(X + X') \\ Y_{in} &= G + j(B + B') \end{aligned} \tag{6.42}$$

It is apparent that R and G, the real parts of Z_{in} and Y_{in}, are entirely unaffected by the presence of arbitrary reactive networks placed as shown in Fig. 6.15. Thus the method of Gewertz can give the complete impedance function from a knowledge of the real part only for "minimum-reactance" or "minimum-susceptance" networks, that is, for networks such as those of Fig. 6.15 in which $X' = 0$ or $B' = 0$.

If an impedance function of p contains more zeros than poles, the denominator may be divided into the numerator to give a sum of two terms, one of which represents a series inductance if the expression is for an impedance or a shunt capacitance if the expression is for an admittance. These elements represent "excess reactance" or "excess susceptance" respectively. Consequently, when applying the method of Gewertz, the degree of the numerator of the immittance expression is never assumed greater than that of the denominator. For the example we treated, this means that a_3 should be assumed to be zero. Then, a unique minimum-reactance impedance can be obtained given only the real part of the impedance.

6.5 Synthesis for an impedanceless source

Consider the network of Fig. 6.16. It is one driven with an impedanceless voltage source and terminated in a single resistance. The transfer function is $F(j\omega) = e_n/e_0$.

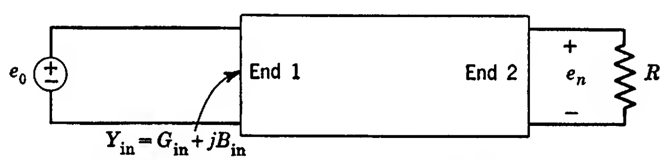


Fig. 6.16. Network with ideal voltage generator.

The value of $|e_n|^2$ is

$$|e_n|^2 = FF^*|e_0|^2 = F(j\omega)F(-j\omega)|e_0|^2 \tag{6.43}$$

The power in the load R is

$$P_n = |e_n|^2/R \tag{6.44}$$

and that furnished by the voltage generator is

$$P_{\text{in}} = G_{\text{in}}|e_0|^2 = \text{Re}\,(Y_{\text{in}})|e_0|^2 \tag{6.45}$$

Since the network is dissipationless, *the power furnished by the voltage source must equal that dissipated in the load resistor.* Hence, we may equate eqs. 6.44 and 6.45 and substitute for $|e_n|^2$ from eq. 6.43 to obtain

$$\text{Re}\,(Y_{\text{in}}) = \text{Re}\,(1/Z_{\text{in}}) = FF^*/R \tag{6.46}$$

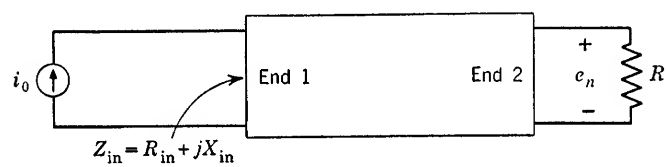
which is the equation of importance here. We design such networks by first specifying the transfer function F. Then, the function FF^*/R having quadrantal symmetry is equated to the real part of the input admittance. The method of Gewertz is then employed in order to find

the complete expression for Y_{in} or Z_{in}. Finally, the network is synthesized as the input immittance of a network, as has been discussed. Since the poles of FF^* are generally known as part of the function design process, those of Y_{in} are specified. It is only the zeros of Y_{in} that need be found.

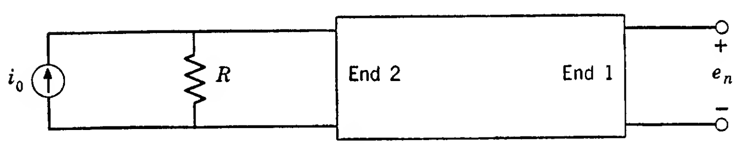
The only word of caution is that FF^* cannot be greater than unity at $\omega = 0$, as should be obvious to the reader. In fact, FF^* will be either unity or zero at $\omega = 0$ and nothing else, because inductors become short circuits and capacitors open circuits.

Any susceptance in shunt with the ideal voltage generator of Fig. 6.16 will have no effect upon the transfer function; hence, it is only the minimum-susceptance network that is of interest as given by Gewertz's method.

A similar procedure applies when an ideal current generator drives the network as in Fig. 6.17a. The solution for the transfer function of



(a)



(b)

Fig. 6.17. Network with ideal current generator.

the network of Fig. 6.17a also applies to the network turned end for end as in Fig. 6.17b. In this case, let the transfer function be a transfer impedance defined by $F(j\omega) = e_n/i_0$. Then

$$|e_n|^2 = FF^*|i_0|^2 \tag{6.47}$$

The power in the load R_L is

$$P_n = |e_n|^2/R \tag{6.48}$$

and that furnished by the current generator is

$$P_{\text{in}} = R_{\text{in}}|i_0|^2 = \text{Re}\,(Z_{\text{in}})|i_0|^2 \tag{6.49}$$

Since the network is dissipationless, P_n and P_{in} must be equal. Thus, equating eqs. 6.48 and 6.49 with the aid of eq. 6.47

$$\text{Re}\,(Z_{in}) = \text{Re}\,(1/Y_{in}) = FF^*/R \tag{6.50}$$

As in the case when the source is a voltage generator, FF^* must be equal to either R^2 or zero at $\omega = 0$.

If the current i_0 is given by $i_0 = ge_0$ where g is finite, and the transfer function is $G = e_n/e_0$, the relation becomes

$$\text{Re}\,(Z_{in}) = \text{Re}\,(1/Y_{in}) = GG^*/g^2R \tag{6.51}$$

Since a reactance in series with the current generator will have no bearing on the transfer function, it is only the minimum-reactance network that is important.

6.6 The general Darlington network

Consider Fig. 6.18*a*, which shows a source and source impedance and a dissipationless network terminated in a resistance. The transfer function can be defined as

$$e_2/e_1 = F(j\omega) \tag{6.52}$$

which is also the transfer function of the network of Fig. 6.18*b* except for a constant multiplier given by a relation containing R_1 and R_2.

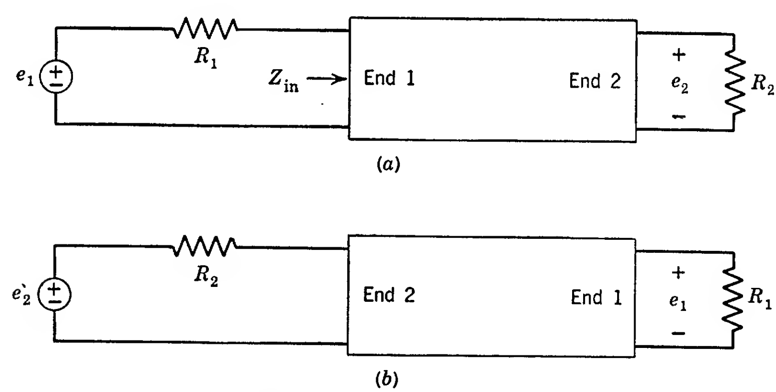


Fig. 6.18. Darlington network with resistive terminations.

The maximum possible power into the network from the voltage source e_1 for the circuit of Fig. 6.18*a* is given when Z_{in} is the complex conjugate of the source resistance R_1. Then

$$(P_1)_{max} = |e_1|^2/4R_1 \tag{6.53}$$

in which effective values have been assumed. Peak values could also be assumed without any change in the final results. The power given by eq. 6.53 is often referred to as the "available" power.

The power P_2 in the load resistor R_2 is the same as that into the network P_1, because the network is dissipationless. Thus, the maximum possible power in R_2 is

$$(P_2)_{\max} = |e_2|^2/R_2 = |e_1|^2/4R_1 \tag{6.54}$$

Placing eq. 6.52 in eq. 6.54, the maximum permissible value of FF^* is obtained as

$$(FF^*)_{\max} = R_2/4R_1 \tag{6.55}$$

At $\omega = 0$, coils are short circuits and capacitors are open circuits; hence, $F(j\omega)$ can be either zero or $R_2/(R_1 + R_2)$ at $\omega = 0$. This, along with eq. 6.55, must be borne in mind when the transfer function obtained from some function design is assigned a constant multiplier.

The ratio R_2/R_1 is an impedance ratio. In part, the network acts to transform a signal from the impedance level R_1 to R_2. In band-pass circuits, the transformation can often be made as good as that obtainable with an ideal transformer without the need for coupled coils. A perfect match can always be obtained if ideal transformers are allowed. Often, however, a conjugate match is not desired. If $R_1 \neq R_2$, a conjugate match cannot be obtained at $\omega = 0$ without an ideal transformer.

The power entering the network from the generator is

$$P_1 = \frac{|e_1|^2 R_{\text{in}}}{|R_1 + Z_{\text{in}}|^2} = P_2 = |e_2|^2/R_2 \tag{6.56}$$

in which the power fed into the network must be equal to the power in the load resistor because the network is lossless. If eq. 6.56 is normalized to the maximum possible power as given by eq. 6.54, we get

$$1 - \frac{P_1}{(P_1)_{\max}} = 1 - \frac{4R_1 R_{\text{in}}}{|R_1 + Z_{\text{in}}|^2} \tag{6.57}$$

Expressing Z_{in} as $R_{\text{in}} + jX_{\text{in}}$ and manipulating

$$1 - \frac{P_1}{(P_1)_{\max}} = \left|\frac{R_1 - Z_{\text{in}}}{R_1 + Z_{\text{in}}}\right|^2 = |\rho|^2 \tag{6.58}$$

where ρ is termed the "reflection coefficient" and $|\rho|$ is bounded by zero and unity. (The synthesis procedure we are working out expresses ρ in terms of the transfer function and then the input impedance is expressed in terms of ρ.)

We thus have

$$P_1 = P_2 = (P_1)_{\max}(1 - |\rho|^2) \tag{6.59}$$

Substituting for P_2 and $(P_1)_{\max}$, eq. 6.59 becomes

$$|e_2/e_1|^2 = F(j\omega)F(-j\omega) = (R_2/4R_1)(1 - |\rho|^2) \tag{6.60}$$

which is the square of the magnitude of the transfer function and which must show quadrantal symmetry. Thus

$$F(p)F(-p) = (R_2/4R_1)[1 - \rho(p)\rho(-p)] \tag{6.61}$$

or

$$\begin{aligned} \rho(p)\rho(-p) &= 1 - \frac{4R_1F(p)F(-p)}{R_2} \\ &= \frac{[R_1 - Z_{\text{in}}(p)][R_1 - Z_{\text{in}}(-p)]}{[R_1 + Z_{\text{in}}(p)][R_1 + Z_{\text{in}}(-p)]} \end{aligned} \tag{6.62}$$

where the second relation has been deduced from eq. 6.58.

Since the transfer function $F(p)$ is known, the poles and zeros of $\rho(p)\rho(-p)$ are known. $Z_{\text{in}}(p)$ is positive real and has p-z only in the left half-plane; hence, $R_1 + Z_{\text{in}}(p)$ has p-z only in the left half-plane. The poles of $\rho(p)\rho(-p)$ that lie in the left half-plane are the zeros of $R_1 + Z_{\text{in}}(p)$, and their negatives in the right half-plane are the zeros of $R_1 + Z_{\text{in}}(-p)$. In order to see this a little more clearly, note that the denominators of $Z_{\text{in}}(p)$ and $Z_{\text{in}}(-p)$ cancel out when expressing the right side of eq. 6.62 as a rational fraction. The poles of $\rho(p)$ are taken in the left half-plane like any stable transfer function. Thus, the poles of $\rho(p)$ are the zeros of $R_1 + Z_{\text{in}}(p)$.

The zeros of $\rho(p)\rho(-p)$ display quadrantal symmetry. The zeros of $\rho(p)$ are chosen from this grouping; they are chosen in complex-conjugate pairs and either a zero or its negative but not both is chosen as that belonging to $\rho(p)$. If the choice exists, more than one network realization will exist.

Finally, when $\rho(p)$ is found, either the positive or negative square root of the multiplier can be taken, which are alternatives leading to the dual networks.

The general form of eq. 6.61 is

$$\rho(p)\rho(-p) = H^2 \frac{(p - z_1)(p + z_1)(p - z_2)(p + z_2)\cdots}{(p - p_1)(p + p_1)(p - p_2)(p + p_2)\cdots} \tag{6.63}$$

from which

$$\rho(p) = \frac{N}{D} = \pm H \frac{(p \pm z_1)(p \pm z_2)(p \pm z_3)\cdots}{(p - p_1)(p - p_2)(p - p_3)\cdots} \tag{6.64}$$

where the factors $(p - p_1)$, $(p - p_2)$ are factors relating to poles in the left half-plane, and $(p \pm z_1)$, $(p \pm z_2)$ refer to a zero z_i or its negative $-z_i$. Every sign in eq. 6.64 indicates a choice.

The reflection coefficient has been given as

$$\rho(p) = \frac{N}{D} = \frac{R_1 - Z_{\text{in}}(p)}{R_1 + Z_{\text{in}}(p)} \tag{6.65}$$

If we express $\rho(p)$ in terms of its numerator and denominator polynomials according to eq. 6.64, we get from eq. 6.65 the final relation

$$Z_{\text{in}}(p) = R_1 \frac{D - N}{D + N} \tag{6.66}$$

From this point on, the network is synthesized as the input impedance of a network as described earlier in the chapter.

If all the zeros of the transfer function $F(p)$ are on the $j\omega$ axis (if the function has zeros), the network can be synthesized as a ladder having L-C shunt and series arms. If $F(p)$ is an all-pole function, the ladder will have a single L in each series arm and a single C in each shunt arm and can easily be obtained with the special L-C ladder synthesis described before.

6.7 Example of synthesis for an impedanceless source

To demonstrate procedures employed in practical synthesis, we shall consider the example of a three-pole function often used before. This example is simple enough to be treated with reasonable ease but is not so simple that it is trivial.

Suppose it is desired to have an ideal current generator as in Fig. 6.17*a*. The value of FF^* at $\omega = 0$ is R^2. Hence

$$\frac{e_2}{i_0} = F(p) = \frac{b_0 R}{p^3 + b_2 p^2 + b_1 p + b_0} \tag{6.67}$$

Then

$$\begin{aligned}\frac{F(p)F(-p)}{R} &= \text{Ev } Z_{\text{in}}(p) \\ &= \frac{b_0^2 R}{(p^3 + b_2 p^2 + b_1 p + b_0)(-p^3 + b_2 p^2 - b_1 p + b_0)}\end{aligned} \tag{6.68}$$

The expression for $Z_{\text{in}}(p)$ is

$$Z_{\text{in}}(p) = \frac{a_3 p^3 + a_2 p^2 + a_1 p + a_0}{p^3 + b_2 p^2 + b_1 p + b_0} \tag{6.69}$$

and that for Ev $Z_{\text{in}}(p)$ is

$$\text{Ev } Z_{\text{in}}(p) = \frac{c_6p^6 + c_4p^4 + c_2p^2 + c_0}{d_6p^6 + d_4p^4 + d_2p^2 + d_0} \tag{6.70}$$

Since it is assumed that the transfer function is given data, we know the values of the c's and d's. In particular,

$$\begin{aligned} c_6 &= c_4 = c_2 = 0 \\ c_0 &= b_0^2R \end{aligned} \tag{6.71}$$

Referring to Sec. 6.4 on the relationship between the real and imaginary parts of input immittances, noting the coefficients that are zero, and with $b_3 = 1$

$$\begin{aligned} c_0 &= b_0^2R = a_0b_0 \\ c_2 &= 0 = a_0b_2 + b_0a_2 - a_1b_1 \\ &= Rb_0b_2 + b_0a_2 - a_1b_1 \\ c_4 &= 0 = a_0b_4 + a_2b_2 + a_4b_0 - a_1b_3 - a_3b_1 \\ &= a_2b_2 - a_1 - a_3b_1 \\ c_6 &= 0 = a_0b_6 + a_2b_4 + a_4b_2 + a_6b_0 - a_1b_5 - a_3b_3 - a_5b_1 \\ &= -a_3b_3 = -a_3 \end{aligned} \tag{6.72}$$

From the last expression of eqs. 6.72, $a_3 = 0$. Then between the expressions for c_2 and c_4

$$a_2 = \frac{Rb_2b_0}{b_1b_2 - b_0} \qquad a_1 = \frac{Rb_2^2b_0}{b_1b_2 - b_0} \qquad a_0 = b_0R \tag{6.73}$$

Thus we have

$$Z_{\text{in}}(p) = R\frac{[(b_2b_0)/(b_1b_2 - b_0)]p^2 + [(b_2^2b_0)/(b_1b_2 - b_0)]p + b_0}{p^3 + b_2p^2 + b_1p + b_0} \tag{6.74}$$

which can be realized with the circuit of Fig. 6.19.

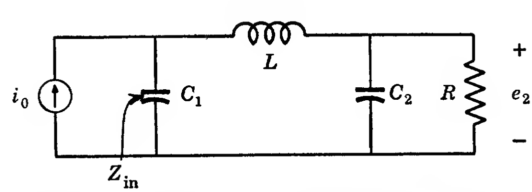


Fig. 6.19. Synthesis for an impedanceless source.

In order to be specific, assume that $b_0 = 1$, $b_1 = b_2 = 2$, and $R = 1$ to give the maximally flat three-pole function. Then

$$Z_{\text{in}}(p) = \frac{(\frac{2}{3})p^2 + (\frac{4}{3})p + 1}{p^3 + 2p^2 + 2p + 1} \tag{6.75}$$

This function has a zero at infinity. Thus, $Y_{\text{in}}(p)$ has a pole on the $j\omega$ axis at infinity and

$$Y_{\text{in}}(p) = (\tfrac{3}{2})p + \frac{(\frac{1}{2})p + 1}{(\frac{2}{3})p^2 + (\frac{4}{3})p + 1} \tag{6.76}$$

The remainder has a zero at infinity; thus, the reciprocal of the remainder has a pole at infinity and

$$Y_{\text{in}}(p) = (\tfrac{3}{2})p + \frac{1}{(\frac{4}{3})p + [1/(\frac{1}{2}p + 1)]} \tag{6.77}$$

Thus, the element values of Fig. 6.19 are $C_1 = \frac{3}{2}$, $C_2 = \frac{1}{2}$, and $L = \frac{4}{3}$.

6.8 Example of the general Darlington method

In order to make comparison with the network of Sec. 6.7, we shall use a similar transfer function. Let us take the transfer function of the network of Fig. 6.18*a* as

$$\frac{e_2}{e_1} = F(p) = \frac{R_2/(R_1 + R_2)}{p^3 + 2p^2 + 2p + 1} \tag{6.78}$$

The square of the reflection coefficient is

$$\begin{aligned} \rho(p)\rho(-p) &= 1 - (4R_1/R_2)F(p)F(-p) \\ &= 1 - \frac{4R_1R_2/(R_1 + R_2)^2}{-p^6 + 1} = 1 - \frac{T}{-p^6 + 1} \end{aligned} \tag{6.79}$$

where T is an impedance parameter defined by eq. 6.79 which is always equal to or less than unity. For simplicity, let us assume for the moment that $R_1 = R_2$. Then, $T = 1$ and

$$\begin{aligned} \rho(p)\rho(-p) &= \frac{p^6}{p^6 - 1} \\ &= \frac{(p + 0)^3(p - 0)^3}{(p^3 + 2p^2 + 2p + 1)(p^3 - 2p^2 + 2p - 1)} \end{aligned} \tag{6.80}$$

The poles of eq. 6.80 lying in the left half-plane belong to $\rho(p)$. Of the six zeros, three belong to $\rho(p)$ and three to $\rho(-p)$. Since the zeros are

all at the origin, no choice is given. Thus

$$\rho(p) = \frac{N}{D} = \frac{\pm p^3}{p^3 + 2p^2 + 2p + 1} \tag{6.81}$$

and

$$\frac{Z_{\text{in}}(p)}{R_1} = \frac{D - N}{D + N} = \frac{p^3 + 2p^2 + 2p + 1 \mp p^3}{p^3 + 2p^2 + 2p + 1 \pm p^3} \tag{6.82}$$

The choice of sign merely inverts the rational function. The two functions for the input impedance, assuming $R_1 = 1$ for simplicity, are

$$Z_{\text{in}}(p) \text{ or } Y_{\text{in}}(p) = \frac{2p^2 + 2p + 1}{2p^3 + 2p^2 + 2p + 1} \tag{6.83}$$

Expanding into continued fractions, we get the circuit of Fig. 6.20*a* when eq. 6.83 is equal to $Z_{\text{in}}(p)$ and the circuit of Fig. 6.20*b* when eq. 6.83 is equal to $Y_{\text{in}}(p)$. The two circuits are exact duals.

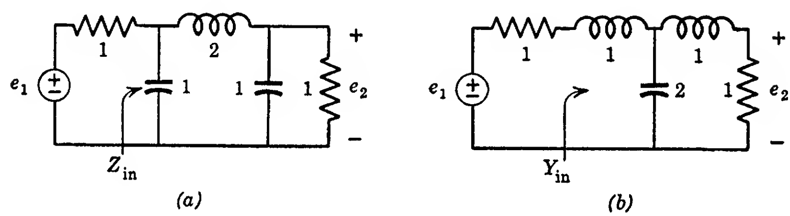


Fig. 6.20. Synthesis with source and load impedances.

Let us now consider what happens when T is not equal to unity. Then

$$\rho(p)\rho(-p) = \frac{p^6 - (1 - T)}{p^6 - 1} \tag{6.84}$$

There are a total of six zeros to this function with the three in the right half-plane being the negatives of the three in the left half-plane. The zeros lie on a circle of radius $(1 - T)^{1/6} = r$ and have equal angles between them. There are a total of eight choices for ρ as

$$\rho(p) = \frac{\pm(p \pm r)(p^2 \pm rp + r^2)}{p^3 + 2p^2 + 2p + 1} \tag{6.85}$$

The $\pm$ signs in front give the dual networks. For each of the two duals, there are four different functions of ρ. The function T is symmetric in R_1 and R_2. Thus, half the values of $\rho(p)$ go with a ratio of R_1/R_2 of, say, k and the other half with R_2/R_1 equal to the same k.

6.9 The general maximally flat all-pole filter

The maximally flat function has been and will continue to be employed in examples throughout the book. This particular function is used so much because it is a fairly sophisticated function yet its mathematics are quite simple. Further, as a function, it lies roughly in the middle of all functions (having only poles) that approximate the brick-wall characteristics. Whereas moving maximally flat poles to the right results in high-efficiency Chebyshev functions, moving them left results in linear-phase functions.

Because the mathematics are so simple, we can study the maximally flat all-pole ladder filter in considerable detail, and in particular can determine the input admittance of the relevant Darlington networks for an arbitrary number of poles and for any source to load resistance ratio.

Let us define the general maximally flat transfer function with *unity* half-power bandwidth as an extension of eq. 6.78 as

$$F(p) = \frac{R_2/(R_1 + R_2)}{p^2 + b_{n-1}p^{n-1} + b_{n-2}p^{n-2} + \cdots + b_1 p + 1} \tag{6.86}$$

The coefficients b_j can be found from the dimensions of the poles on the unit circle to any accuracy that is desired, or from the general expressions given in Chap. 5.

Following eq. 6.84, the square of the reflection coefficient is

$$\rho(p)\rho(-p) = 1 - (4R_1/R_2)F(p)F(-p) = \frac{\pm p^{2n} + (1 - T)}{\pm p^{2n} + 1} \tag{6.87}$$

where the positive sign is for n even and the negative sign for n odd and where

$$T = \frac{4R_1R_2}{(R_1 + R_2)^2} \tag{6.88}$$

The zeros of $\rho(p)$ lie on a circle of radius $(1 - T)^{1/2n}$. The poles of $\rho(p)$ lie in the left half-plane on a circle of unit radius. We may therefore express $\rho(p)$ as

$$\rho(p) = \pm\frac{p^n + a_{n-1}p^{n-1} + \cdots + a_2p^2 + a_1p \pm (1 - T)^{1/2}}{p^n + b_{n-1}p^{n-1} + \cdots + b_2p^2 + b_1p + 1} = \frac{N}{D} \tag{6.89}$$

where the $\pm$ multiplier on $(1 - T)^{1/2}$ depends upon how we take the zeros of $\rho(p)$.

Let us now (and henceforth in this section) stipulate that the resulting ladder network must have a shunt capacitance next to the source resistance R_1. (The dual network will then have a series inductance.)

Then, the expression for the input admittance must have a numerator of higher degree in p than the denominator. Therefore in finding the input impedance from

$$Z_{\text{in}}(p) = R_1 \frac{D + N}{D - N} \tag{6.90}$$

the positive multiplier of eq. 6.89 must be used to give

$$Y_{\text{in}} = G_1 \frac{2p^n + (b_{n-1} + a_{n-1})p^{n-1} + \cdots + (b_1 + a_1)p + (1 \pm \sqrt{1 - T})}{(b_{n-1} - a_{n-1})p^{n-1} + \cdots + (b_1 - a_1)p + (1 \mp \sqrt{1 - T})} \tag{6.91}$$

The leading capacitance is evidently

$$C_1 = \frac{2G_1}{b_{n-1} - a_{n-1}} \tag{6.92}$$

The coefficient b_{n-1} is the negative of the sum of the pole positions of $\rho(p)$. The coefficient a_{n-1} is the negative of the sum of the zero positions of $\rho(p)$. Clearly, a_{n-1} is a maximum if the zeros of $\rho(p)$, like the poles, are all taken in the left half-plane. This choice maximizes C_1, which permits the source to drive the largest possible capacitance. Throughout the balance of this section, the leading capacitance will be assumed to be maximized in this manner.

Since both the p-z of $\rho(p)$ lie on circles with equal angular spacing, and because the number of zeros of $\rho(p)$ is equal to the number of poles, the coefficients a_i can be expressed in terms of the b_i as

$$\begin{aligned} a_{n-1} &= b_{n-1}(1 - T)^{1/2n} \\ a_{n-2} &= b_{n-2}(1 - T)^{2/2n} \\ &\cdot \\ &\cdot \\ &\cdot \\ a_{n-k} &= b_{n-k}(1 - T)^{k/2n} \end{aligned} \tag{6.93}$$

Thus, the general expression for the input admittance of the n-pole maximally flat Darlington network operating between source and load resistances R_1 and R_2 respectively, and having a maximum input capacitance, is

$$Y_{\text{in}} = G_1 \frac{2p^n + b_{n-1}[1 + (1 - T)^{1/2n}]p^{n-1} + \cdots + [1 + (1 - T)^{1/2}]}{b_{n-1}[1 - (1 - T)^{1/2n}]p^{n-1} + \cdots + [1 - (1 - T)^{1/2}]} \tag{6.94}$$

The coefficients b_j can be determined from the pole positions of $F(p)$ (the transfer function). The parameter T depends upon the ratio of R_1 and R_2. Knowing R_1, R_2, and n, the admittance can be expressed in numerical terms and the network determined by means of the special synthesis derived earlier in the chapter. It should be noted that a fair degree of numerical accuracy may be required in the calculations.

The work of finding a network and the loss in numerical accuracy which results after several steps in the continued-fraction expansion (or the special procedure described in Sec. 6.3) can often be reduced by working from both ends of the filter. In order to do this, the filter can be designed not only for the given values of R_1 and R_2 but also for the reversed case where R_2 becomes the source and R_1 the load. In addition, the dual networks can be found. Between the expressions for the input impedance and admittance for R_1 the source and comparable expressions for R_2 the source, continued-fraction expansions will lead to the determination of the element values expanding from both ends of the filter towards the middle.

Special cases of eq. 6.94 are of interest. First is the matched filter where $R_1 = R_2 = R$. Then, $T = 1$ and eq. 6.94 simplifies to

$$Y_{\text{in}} = G\frac{2p^n + b_{n-1}p^{n-1} + \cdots + b_1p + 1}{b_{n-1}p^{n-1} + \cdots + b_1p + 1} \tag{6.95}$$

The matched filter is physically symmetric for n odd, which makes it the easiest to find from the continued-fraction expansion.

The other special case of interest is a limiting one where $G_1 \to 0$, which corresponds to the ideal current source. To determine the input admittance for this case, limiting expressions must be obtained as

$$\lim_{R_1 \to \infty} R_1 b_{n-k}[1 - (1 - T)^{k/2n}] = 2kR_2/n \tag{6.96}$$

which gives

$$\lim_{R_1 \to \infty} Y_{\text{in}} = G_2 \frac{p^n + b_{n-1}p^{n-1} + \cdots + b_1p + 1}{\dfrac{b_{n-1}}{n}p^{n-1} + \dfrac{2b_{n-2}}{n}p^{n-2} + \cdots + \dfrac{(n-1)b_1}{n}p + 1} \tag{6.97}$$

The leading capacitance of the filter for any source to load resistance ratio is found from eq. 6.92 as

$$C_1 = \frac{2G_1}{b_{n-1}[1 - (1 - T)^{1/2n}]} \tag{6.98}$$

and

$$\lim_{R \to \infty} R_2C_1 = n/b_{n-1} \tag{6.99}$$

If we had gone through the same argument in order to obtain the minimum value of C_1 rather than the maximum value, we would have chosen all the zeros of the reflection coefficient in the right half-plane to get

$$C_n = \frac{2G_2}{b_{n-1} + a_{n-1}} = \frac{2G_2}{b_{n-1}[1 + (1 - T)^{1/2n}]} \tag{6.100}$$

and

$$\lim_{R_1 \to \infty} R_2 C_n = 1/b_{n-1} \tag{6.101}$$

One choice of the zeros of $\rho(p)$ applies to a resistance ratio R_1/R_2 equal to some number, whereas the ratio R_2/R_1 equal to the same number applies to exactly the opposite choice for the zeros. Therefore, for n odd and for $R_1 > R_2$, eq. 6.100 is that for the capacitance adjacent to the load R_2. The capacitance ratio (input capacitance to output capacitance) is

$$\frac{C_1}{C_n} = \frac{R_2}{R_1}\left(\frac{1 + (1 - T)^{1/2n}}{1 - (1 - T)^{1/2n}}\right) \tag{6.102}$$

which is unity for the matched case but which departs rapidly from unity as T differs from unity. In the limit as $R_1 \to \infty$, it becomes

$$\lim_{R_1 \to \infty} \frac{C_1}{C_n} = n \tag{6.103}$$

The rapid change in C_1 with R_1/R_2 implies that with but a small sacrifice in power transfer at $\omega = 0$ (that is, with T only slightly less than unity) a much larger value of C_1 can be tolerated than that applicable to the matched system. This has important advantages when it is desired to build a filter with the maximum possible bandwidth and with a relatively large gain when constrained by some unavoidable shunt capacitance. In fact, we shall find (using a rather meaningful criterion) that a surprisingly large intentional mismatch gives a big improvement.

The power gain of the two-terminal maximally flat low-pass filter at $\omega = 0$ relative to the maximum possible power gain (when $R_1 = R_2$) is simply T. If the power gain transfer function relative to the matched case is plotted, there results a curve like that of Fig. 6.21. Let us define a tolerance bandwidth B where the relative power gain is Q as shown in Fig. 6.21 (which is not the usual half-power bandwidth). Since

$$\text{Relative power gain} = \frac{T}{1 + \omega^{2n}} \tag{6.104}$$

we find

$$B = (T/Q - 1)^{1/2n}, \quad T \geq Q \tag{6.105}$$

Let us form the product of tolerance bandwidth, leading capacitance C_1, source resistance R_1, and tolerance value Q as

$$R_1C_1BQ = \frac{2Q(T/Q - 1)^{1/2n}}{b_{n-1}[1 - (1 - T)^{1/2n}]}, \quad T \geq Q \tag{6.106}$$

The curve of this function has a rather broad maximum at some value of T other than unity. The value of T at this maximum obviously gives the optimum mismatch required for maximizing the product of

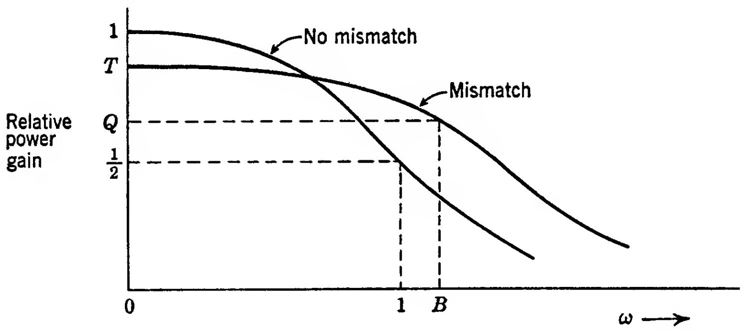


Fig. 6.21. Relative power gain with and without mismatch (for constant R_1C_1).

eq. 6.106, which is the *area* BQ of Fig. 6.21 for a fixed R_1C_1. Clearly, since C_1 is larger for $R_1 > R_2$ than for $R_1 < R_2$, a larger source resistance than load resistance is required.

The maximum of eq. 6.106 can be established in the usual manner by setting its derivative with respect to T equal to zero. There results

$$T_0 = 1 - (1 - Q)^{2n/(2n-1)} \tag{6.107}$$

where T_0 is the optimum value of T.

As a specific numerical example, assume $n = 3$ and $Q = \frac{1}{2}$. Then, eq. 6.107 gives $T_0 = 0.565$. By solving eq. 6.88 for R_1/R_2, we get

$$R_1/R_2 = 2/T - 1 + (2/T)(1 - T)^{1/2} \tag{6.108}$$

Using $T_0 = 0.565$ in this equation, the optimum resistance ratio is determined as $R_1/R_2 = 4.87$. This is indeed a surprisingly large mismatch. From eq. 6.105, the tolerance bandwidth for this example is calculated as 0.711. The product R_1C_1 is found to be 7.7. Thus, the product of interest is $R_1C_1BQ = 2.74$ when optimum mismatch is employed. It is of interest to compare this figure to that pertaining to the matched filter. For the matched filter $R_1C_1 = 1$ and the tolerance bandwidth is unity. Therefore, $R_1C_1BQ = 0.5$. The impressive advantage gained through intentional mismatch should be clearly understood.

Problems

1. Using a continued-fraction expansion, find the Darlington network represented by the input immittance

$$3 \times 10^{-9} \frac{p^3 + 2 \times 10^6 p^2 + 2 \times 10^{12} p + 10^{18}}{p^2 + 2 \times 10^6 p + 1.5 \times 10^{12}}$$

2. Write the input admittance for the circuit of Fig. P.2. Then, work out a procedure using this circuit as an example so that you will be able to synthesize circuits of this general type given only an input immittance.

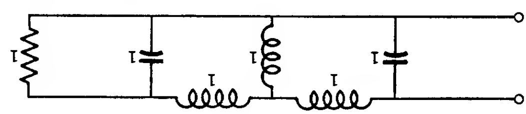


Fig. P.2.

3. Obtain the Darlington network for the following immittance in which the transfer function from which it was derived has zeros on the $j\omega$ axis

$$1.5 \times 10^{-6} \frac{p^3 + 667p^2 + 667 \times 10^5 p + 333 \times 10^8}{p^2 + 500p + 500 \times 10^5}$$

4. Find the Darlington network represented by the following input immittance in which a combination of partial- and continued-fraction expansions should be used

$$\frac{p^6 + p^5 + 3p^4 + 3p^2 + 2p + 1}{p^5 + p^4 + p^3 + p^2 + p}$$

5. An input impedance has a real part $1/(1 + \omega^6)$. Find the complete expression for the impedance and obtain the network.

6. The real part of an input impedance is

$$\frac{500}{1 + 10^{-14}\omega^2 + 10^{-28}\omega^4}$$

Find the expression for the impedance and determine the network.

7. It is desired to have a transfer function with three poles displaced as a Q-D function and a pair of zeros on the $j\omega$ axis to give infinite rejection at one frequency. The function is

$$H \frac{T^2p^2/12 + 1}{1 + Tp + T^2p^2/2 + T^3p^3/6}$$

The network is to work between 1 Ω source and load impedances. Determine the Darlington network and plot the locations of the p-z. Any convenient frequency normalization may be used.

8. Find the networks for $R_1 = 2R_2$ for the function treated as an example in Sec. 6.8.

9. Find the networks for $R_1 = 4R_2$ for the function treated as an example in Sec. 6.8.

10. The network of Fig. 6.20a is maximally flat and has a bandwidth of unity. Convert this network to a band-pass network at a center frequency of ω_0. Then, using determinant manipulation, make the load impedance (a) as large as is possible

and (*b*) as small as is possible without affecting the input impedance. Will there exist a conjugate match after determinant manipulation?

11. Determine the coefficients b_{n-1}, b_{n-2}, $\cdots$ for a maximally flat all-pole function having unity half-power bandwidth for $n = 2, 3, 4,$ and 5. Find these coefficients using five-place logarithms.

12. Find expressions for the input admittances of the filters of Fig. P.12 in which C_1 is a maximum and the transfer function is maximally flat with unity half-power bandwidth.

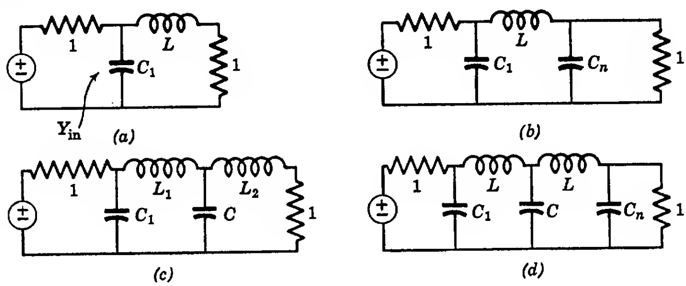


Fig. P.12.

13. Find all element values for the circuits of Fig. P.12.

14. Repeat Prob. 12 for the circuits of Fig. P.14.

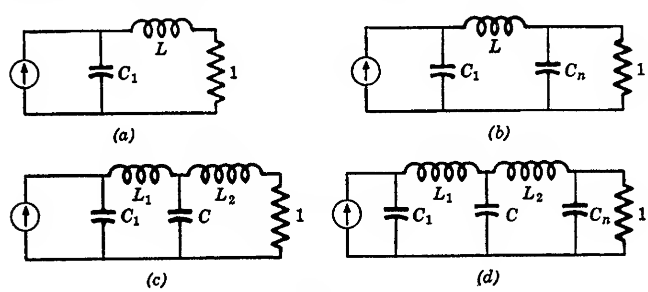


Fig. P.14.

15. Repeat Prob. 13 for the circuits of Fig. P.14.

16. Derive eq. 6.96.

17. The product R_1C_1BQ as given by eq. 6.106 can be maximized, which gives eq. 6.107 from which T_0 can be calculated. If importance is affixed to a large value of relative power gain at $\omega = 0$, then a maximization of the product R_1C_1BQT is perhaps more significant. Derive the relation analogous to eq. 6.107 applicable to the case of interest here.

18. With the results of Prob. 17 and for $n = 3$ and $Q = \frac{1}{2}$, find T_0.

19. Derive the optimum three-pole network (that is, determine element values) for $n = 3$ and $Q = \frac{1}{2}$ for the example in the text following eq. 6.108.

20. Repeat Prob. 19 for the network optimized according to Prob. 18 and compare values to those of the network of Prob. 19.

21. Optimization of R_1C_1BQ yields the optimum ratio R_2/R_1 for the maximally flat all-pole filter. How near to optimum would you expect this resistance ratio to be if applied to a Chebyshev or a linear-phase all-pole filter? Discuss.

7

Image Parameters

We shall leave p-z in this chapter to take up the study of filters designed on an image-impedance basis. Such filters are considered classic in the historical sense. This does not mean, however, that they are not useful. When it is desired to build a complicated filter not including vacuum tubes and having an arbitrary number of pass bands and stop bands (where the signal is transmitted and attenuated respectively), it may be best to employ classical methods simply because of the mathematical difficulties attendant with p-z techniques. In this chapter we shall study the basic theory of classical (or conventional) filters and put off until the next chapter the practical and specific networks.

Image-matched filters have two major drawbacks. First, except for constant-resistance networks, the ideal theoretical gain and phase functions are only approximated when practical source and load impedances are utilized. Second, and again except for constant-resistance networks, an approximate conjugate match between source and load is required (for symmetric filters). It is this second limitation as much as the first that causes image-matched filters to be undesirable for many applications, as when a match is not required or is not possible, or when large advantages can be realized by means of an intentional mismatch.

7.1 Specifications of the two-terminal pair

Consider the circuit of Fig. 7.1*a*; it is one driven with a voltage generator. Rather than designate a specific load impedance, we shall merely designate the load voltage e_2. Similarly, we can indicate the voltage e_1 rather than the voltage source. The result of this is the network of Fig. 7.1*b*. Finally, in order to make the terminal conditions completely symmetric, the current i_2 can be assumed to flow opposite to that shown in Fig. 7.1*b* so that we arrive at the circuit of Fig. 7.1*c*, in which the two ends of the network are called 1 and 2. This circuit is "standard" in the study of classical filter theory (as well as in the study of transistors). The symmetric network of Fig. 7.1*c* could be driven with either a voltage or a current generator.

Let us perform a loop analysis of the network of Fig. 7.1*c*. We shall assume that the numerical labelling of the loops is such that loops 1 and 2 are those associated with the two pairs of terminals. Then, noting that the assumed current i_2 is opposite to convention (clockwise) and that e_1 is a voltage rise in the direction of i_1 whereas e_2 is a voltage

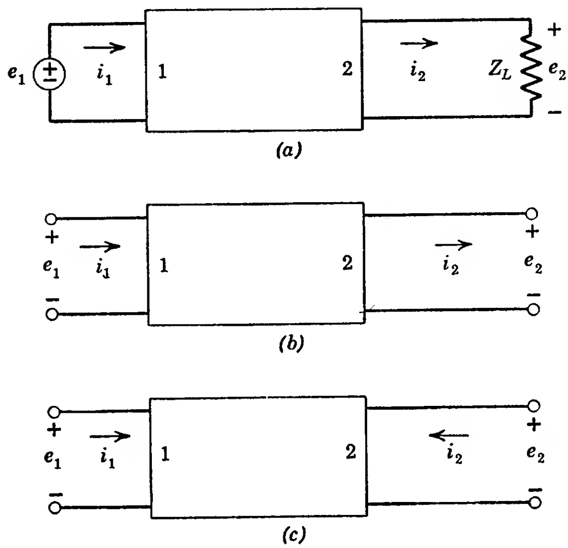


Fig. 7.1. Two-terminal-pair notation.

drop, we get by superposition of the effects resulting from the two voltage generators

$$\begin{aligned} i_1 &= \frac{\Delta_{11}{}^m}{\Delta^m} e_1 - \frac{\Delta_{12}{}^m}{\Delta^m}(-e_2) \\ -i_2 &= -\frac{\Delta_{21}{}^m}{\Delta^m} e_1 + \frac{\Delta_{22}{}^m}{\Delta^m}(-e_2) \end{aligned} \tag{7.1}$$

Let us define new admittances such that eqs. 7.1 can be written

$$\begin{aligned} i_1 &= y_{11}{}^a e_1 + y_{12}{}^a e_2 \\ i_2 &= y_{21}{}^a e_1 + y_{22}{}^a e_2 \end{aligned} \tag{7.2}$$

If vacuum tubes, transistors, or other unilateral devices are present in the network, $y_{21}{}^a$ will not be equal to $y_{12}{}^a$. In this and the following chapter in which classical theory is of interest, we shall always assume

that the networks are bilateral such that these two admittances are identical.

Suppose end 2 of the network of Fig. 7.1c is short-circuited. Then, $e_2 = 0$ and $y_{11}{}^a = i_1/e_1$. We call this a "short-circuit admittance." Thus

$$\begin{aligned} y_{1s} &= y_{11}{}^a = 1/z_{1s} = \text{Input } Y \text{ at end 1 with end 2 short-circuited} \\ y_{2s} &= y_{22}{}^a = 1/z_{2s} = \text{Input } Y \text{ at end 2 with end 1 short-circuited} \end{aligned} \tag{7.3}$$

Now let us make a very similar analysis of the same circuit with node voltages and a nodal analysis. We get

$$\begin{aligned} e_1 &= \frac{\Delta_{11}{}^n}{\Delta^n} i_1 - \frac{\Delta_{12}{}^n}{\Delta^n} i_2 = z_{11}{}^b i_1 + z_{12}{}^b i_2 \\ e_2 &= -\frac{\Delta_{21}{}^n}{\Delta^n} i_1 + \frac{\Delta_{22}{}^n}{\Delta^n} i_2 = z_{21}{}^b i_1 + z_{22}{}^b i_2 \end{aligned} \tag{7.4}$$

in which a new set of z's is defined. If we open-circuit end 2 of the network, $z_{11}{}^b = e_1/i_1$, which we call an "open-circuit impedance." Thus

$$\begin{aligned} z_{10} &= z_{11}{}^b = 1/y_{10} = \text{Input } Z \text{ at end 1 with end 2 open} \\ z_{20} &= z_{22}{}^b = 1/y_{20} = \text{Input } Z \text{ at end 2 with end 1 open} \end{aligned} \tag{7.5}$$

The set of simultaneous equations of eqs. 7.4 can be solved for any pair of the four variables in terms of the remaining two. Of particular interest are

$$\begin{aligned} i_1 &= \frac{z_{22}{}^b}{\Delta_z} e_1 + \frac{-z_{12}{}^b}{\Delta_z} e_2 \\ i_2 &= \frac{-z_{21}{}^b}{\Delta_z} e_1 + \frac{z_{11}{}^b}{\Delta_z} e_2 \end{aligned} \tag{7.6}$$

where $\Delta_z = z_{11}{}^b z_{22}{}^b - z_{12}{}^b z_{21}{}^b$. This set of equations can be equated term by term to eqs. 7.2, which serves to evaluate the y's in terms of the z's. Similarly, we can solve eqs. 7.2 for e_1 and e_2 to obtain

$$\begin{aligned} e_1 &= \frac{y_{22}{}^a}{\Delta_y} i_1 + \frac{-y_{12}{}^a}{\Delta_y} i_2 \\ e_2 &= \frac{-y_{21}{}^a}{\Delta_y} i_1 + \frac{y_{11}{}^a}{\Delta_y} i_2 \end{aligned} \tag{7.7}$$

where $\Delta_y = y_{11}{}^a y_{22}{}^a - y_{12}{}^a y_{21}{}^a$ and which, when compared to eqs. 7.4, serves to evaluate the z's in terms of the y's. We thus have

$$\begin{aligned} y_{11}{}^a &= z_{22}{}^b/\Delta_z & y_{22}{}^a &= z_{11}{}^b/\Delta_z \\ z_{11}{}^b &= y_{22}{}^a/\Delta_y & z_{22}{}^b &= y_{11}{}^a/\Delta_y \\ y_{12}{}^a &= y_{21}{}^a = -z_{12}{}^b/\Delta_z & z_{12}{}^b &= z_{21}{}^b = -y_{12}{}^a/\Delta_y \end{aligned} \tag{7.8}$$

Observe that all of these immittances with the exception of $y_{12}{}^a$ and $z_{12}{}^b$ are simply open- and short-circuit parameters. But, from eqs. 7.8, we can evaluate $y_{12}{}^a$ and $z_{12}{}^b$ in terms of these same open- and short-circuit immittances. It can be concluded that: *The external behavior of a two-terminal-pair network can be completely described in terms of the open- and short-circuit immittances y_{1s}, y_{2s}, z_{10}, and z_{20}.*

Other useful relations growing out of eqs. 7.8 whose derivation is left to the reader as an exercise are

$$\begin{aligned} y_{12}{}^a &= \pm\sqrt{y_{1s}(y_{2s} - y_{20})} = \pm\sqrt{y_{2s}(y_{1s} - y_{10})} \\ z_{12}{}^b &= \pm\sqrt{z_{10}(z_{20} - z_{2s})} = \pm\sqrt{z_{20}(z_{10} - z_{1s})} \end{aligned} \tag{7.9}$$

where the sign is arbitrary; it depends upon the existence of an ideal 1:1 transformer to provide sign inversion between input and output. Also falling out of the manipulations leading to eqs. 7.9 is

$$z_{10}y_{1s} = z_{20}y_{2s} \tag{7.10}$$

which means that only three independent parameters rather than four are needed to describe the behavior of the network.

Another pair of variables useful in the theory of conventional filters can be determined from the preceding relations. If we solve eqs. 7.2 for e_1 and i_1, we get

$$\begin{aligned} e_1 &= \frac{-y_{22}{}^a}{y_{21}{}^a} e_2 - \frac{-1}{y_{21}{}^a} i_2 = Ae_2 - Bi_2 \\ i_1 &= \frac{-\Delta_y}{y_{21}{}^a} e_2 - \frac{-y_{11}{}^a}{y_{21}{}^a} i_2 = Ce_2 - Di_2 \end{aligned} \tag{7.11}$$

where the symbols A, B, C, and D are so defined by almost universal agreement. We may arrive at the same equations starting from eqs.

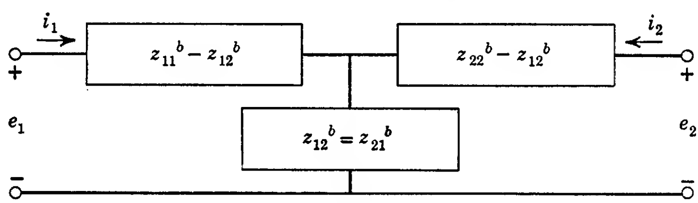
7.4 to get

$$
\begin{aligned}
e_1 &= \frac{z_{11}{}^b}{z_{21}{}^b} e_2 - \frac{\Delta_z}{z_{21}{}^b} i_2 = Ae_2 - Bi_2 \\
i_1 &= \frac{1}{z_{21}{}^b} e_2 - \frac{z_{22}{}^b}{z_{21}{}^b} i_2 = Ce_2 - Di_2
\end{aligned}
\tag{7.12}
$$

Finally, we can solve for the variables e_2 and i_2 in terms of e_1 and i_1 starting from either eqs. 7.2 or 7.4. The result is

$$
\begin{aligned}
e_2 &= De_1 - Bi_1 \\
i_2 &= Ce_1 - Ai_1
\end{aligned}
\tag{7.13}
$$

The manner in which we have specified the terminal conditions of the two-terminal pair is reminiscent of the equations of T and pi networks. It should be a fairly simple matter for the reader to verify the equivalent



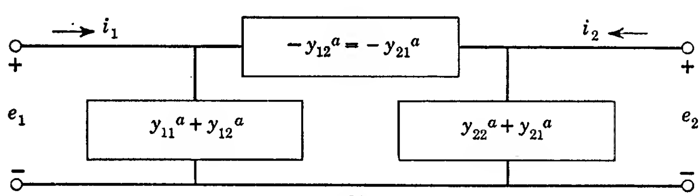


Fig. 7.2. Equivalent pi and T networks.

circuits of Fig. 7.2 for the bilateral network. Of course, the individual immittances of Fig. 7.2 may not correspond to physical bilateral immittances. Nevertheless, these circuits are at least equivalent circuits which, in the steady state, can always be constructed from physical circuit elements to be exact at least at one frequency.

Also of interest is the input immittance to a two-terminal-pair network when terminated in an impedance. If end 2 is terminated in an

impedance Z_2, the input impedance at end 1 is easily found from Fig. 7.2 as

$$
\begin{aligned}
Z_{i1} &= z_{11}{}^b - z_{12}{}^b + \frac{1}{(1/z_{12}{}^b) + [1/(z_{22}{}^b - z_{12}{}^b + Z_2)]} \\
Y_{i1} &= y_{11}{}^a + y_{12}{}^a + \frac{1}{(1/-y_{12}{}^a) + [1/(y_{22}{}^a + y_{21}{}^a + Y_2)]}
\end{aligned}
\tag{7.14}
$$

The impedance looking into end 2, Z_{i2}, with end 1 loaded with an impedance Z_1, can be found from eqs. 7.14 by interchanging the subscripts 1 and 2. With a combination of open- and short-circuit parameters, the expressions for input impedance can be simplified to

$$
\begin{aligned}
Z_{i1} &= z_{10}\frac{Z_2 + z_{2s}}{Z_2 + z_{20}} \\
Z_{i2} &= z_{20}\frac{Z_1 + z_{1s}}{Z_1 + z_{10}}
\end{aligned}
\tag{7.15}
$$

7.2 A simplified development of image-impedance connections

It is desirable when dealing with a chain of filters to be able to obtain the transfer function of the entire filter chain by taking the product of the transfer functions of the individual filter sections. In order to do

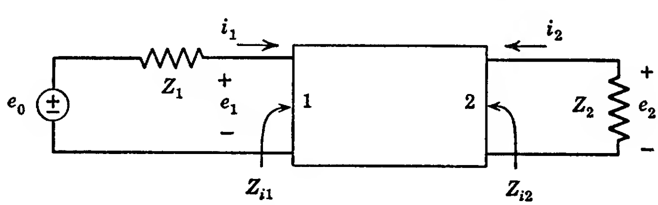


Fig. 7.3. Two-terminal-pair with source and load.

this, the source and load impedances between which each filter section operates must have very specific values.

Consider, for example, the two-terminal-pair network of Fig. 7.3. In order that we be able to find a reasonably simple relation between e_1 and e_0, it is desirable to make the input impedance at end 1 proportional to Z_1. The simplest choice is to make $Z_1 = Z_{i1}$ so that $e_1 = e_0/2$. Although this does not lead to a conjugate match in general, it is hoped that when pure resistances are used for source and load there will exist an approximate conjugate match.

As far as end 2 is concerned, the network and source can, by virtue of Thévenin's theorem, be replaced with a single source voltage and impedance. The impedance of the Thévenin source equivalent is Z_{i2}. Then, similar reasoning applied to end 2 as was applied to end 1 requires that $Z_2 = Z_{i2}$.

If we let $Z_1 = Z_{i1} = z_1$ and $Z_2 = Z_{i2} = z_2$, the network of Fig. 7.3 becomes that of Fig. 7.4 in which the impedances looking either way at the input to the network (and at the output) are the same. Impedances

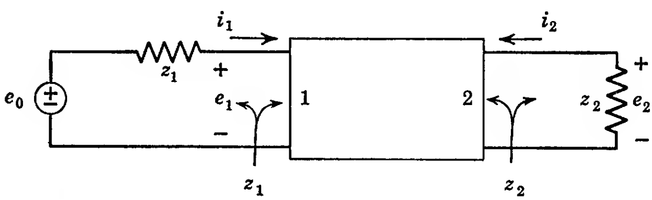


Fig. 7.4. An imaged-matched two-terminal-pair network.

z_1 and z_2 are called the "image" impedances at ends 1 and 2 (input and output) of the network respectively.

If the network is physically symmetric, z_1 and z_2 are equal. Then the source and load impedances must be the same. This means that transforming from one impedance level to another with determinant manipulation or transformers can be effected only after the symmetric network has been designed. The symmetric network is by far the most important conventional filter section.

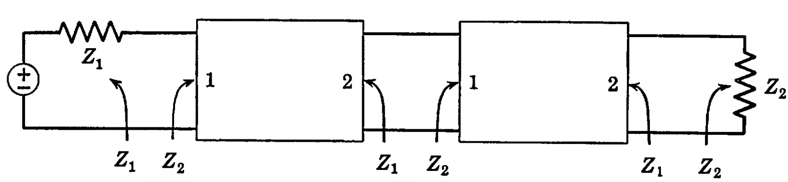


Fig. 7.5. A system matched on the iterative basis.

When the source and load are matched as indicated in Fig. 7.4, the system is said to be matched on the image-impedance basis, the word image being used to distinguish this match of impedances from a complex-conjugate match.

Another matching procedure is to make the impedances looking in each direction along a filter chain equal as is characterized by Fig. 7.5. When a network is matched in this manner, it is said to be matched on an iterative basis. When the network is physically symmetric, Z_1

and Z_2 in Fig. 7.5 will be equal and the iterative and image systems are the same. The unsymmetric iterative system is sometimes applied to resistive attenuators but otherwise is not of such practical value as the image system (although it is gaining in importance in transistor amplifier chains). However, since the symmetric iterative system is the same as the symmetric image system, the iterative notion at least provides a different viewpoint. It can be observed that the unsymmetric iterative system provides little chance for achieving a conjugate match between the source and load.

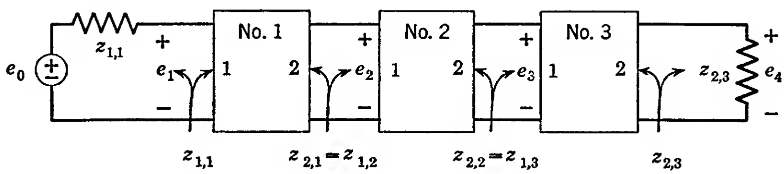


Fig. 7.6. A system matched on the image basis.

Consider now a chain of (three) networks matched on the image basis as in Fig. 7.6. The notation implies that impedance $z_{1,3}$ is the image impedance at end 1 of the third network. Similarly, impedance $z_{2,1}$ is the image impedance at end 2 of the first network. Because of the careful attendance to the image match, each network can be studied by itself because each is terminated in its own image impedance and driven with a source impedance equal to its own image impedance. Thus, we may separately calculate e_2/e_1, e_3/e_2, and e_4/e_3. The product of these three transfer functions gives the over-all transfer function e_4/e_1. Since $e_1 = e_0/2$, we may then easily determine the ratio e_4/e_0.

In studies of conventional filters, it is customary to work with the input-to-output voltage, current, or power ratio rather than with the output-to-input ratio, which has been defined as the transfer function. We shall define the input-to-output ratio as the "reverse" function and thus avoid confusion with the transfer function. Since the reverse and transfer functions are merely reciprocals, the reader should have no real trouble thinking in terms of either. The only thing that may cause confusion is the p-z nomenclature of the function. Obviously, a pole of the transfer function is a zero of the reverse function and conversely. If there is any chance of confusion, the terms "point of infinite loss" and "point of infinite gain" can be used instead of pole and zero.

The question finally arises as to just what makes up these mysterious image impedances. Recall eqs. 7.15, which give the input impedance to a network Z_{i1} when the network is terminated in the impedance Z_2. Setting $Z_1 = Z_{i1} = z_1$ and $Z_2 = Z_{i2} = z_2$ in these equations, we get

$$z_1 = z_{10}\frac{z_2 + z_{2s}}{z_2 + z_{20}}$$
$$z_2 = z_{20}\frac{z_1 + z_{1s}}{z_1 + z_{10}} \tag{7.16}$$

from which we can solve for the image impedances as

$$z_1 = 1/y_1 = (z_{10}z_{1s})^{1/2} = 1/(y_{10}y_{1s})^{1/2}$$
$$z_2 = 1/y_2 = (z_{20}z_{2s})^{1/2} = 1/(y_{20}y_{2s})^{1/2} \tag{7.17}$$

which are the *geometric means of the open- and short-circuit impedances of the network.*

It is now necessary to find the reverse (or transfer) function. We shall use Fig. 7.4 in this discussion. Let us start from the equations

$$e_1 = z_{11}{}^b i_1 + z_{12}{}^b i_2$$
$$e_2 = z_{21}{}^b i_1 + z_{22}{}^b i_2 \tag{7.18}$$

Since $i_1z_1 = e_1$ and $i_2z_2 = -e_2$, these equations can be put in terms of e_1 and e_2 only. We get two equations as

$$\frac{e_1}{e_2} = \frac{z_1}{z_2}\left(\frac{z_{12}{}^b}{z_{11}{}^b - z_1}\right) = \frac{z_1}{z_2}\left(\frac{z_{22}{}^b + z_2}{z_{12}{}^b}\right) \tag{7.19}$$

If each term of eq. 7.19 is used once in finding the square of e_1/e_2, there is obtained

$$\frac{z_2}{z_1}\left(\frac{e_1}{e_2}\right)^2 = (e^{\phi})^2 = e^{2\phi} = e^{2(\alpha+j\beta)} = \frac{z_1}{z_2}\left(\frac{z_{22}{}^b + z_2}{z_{11}{}^b - z_1}\right) \tag{7.20}$$

in which we have defined an exponential reverse function e^{ϕ} which, when squared, becomes $e^{2\phi}$. If we break ϕ into real and imaginary parts $\alpha + j\beta$, then β relates the phase angle of input and output, α relates the magnitude of input and output, and $e^{2\phi} = e^{2\alpha}$ relates the power between input and output.

In terms of open- and short-circuit parameters, eq. 7.20 can be written

$$e^{2\phi} = \sqrt{\frac{z_{1s}z_{20}}{z_{2s}z_{10}}}\left(\frac{1 + \sqrt{z_{2s}/z_{20}}}{1 - \sqrt{z_{1s}/z_{10}}}\right) \tag{7.21}$$

Finally with the aid of eq. 7.10, we get

$$e^{2\phi} = \left(\frac{e_1}{e_2}\sqrt{\frac{z_2}{z_1}}\right)^2 = \left(\frac{i_1}{i_2}\sqrt{\frac{z_1}{z_2}}\right)^2$$
$$= \frac{1 + \sqrt{z_{2s}/z_{20}}}{1 - \sqrt{z_{2s}/z_{20}}} = \frac{1 + \sqrt{z_{1s}/z_{10}}}{1 - \sqrt{z_{1s}/z_{10}}} \tag{7.22}$$

Notice particularly that $e^{2\phi}$ depends upon only two parameters.

7.3 Formal development of image-impedance specifications

In this section, we shall derive the relations obtained in the preceding section in a different manner and one that explains the subject more precisely. More or less, the attempt here is to make the reasoning independent of that in Sec. 7.2 and more complete in many respects.

It is our intention to search for a convenient way to express the transfer function of a two-terminal-pair network. Mainly, this will hinge on the determination of a suitable load impedance such that several filter sections can be cascaded with the over-all reverse function becoming the product of the individual functions. In determining the proper operating conditions, we hope to require knowledge of only the open- and short-circuit parameters; we know them to be sufficient for uniquely specifying the behavior of a bilateral two-terminal pair.

First, let us define the "image impedances" z_1 and z_2 as the geometric means of the open- and short-circuit impedances as

$$\begin{aligned} z_1 &= 1/y_1 = (z_{10}z_{1s})^{1/2} = 1/(y_{10}y_{1s})^{1/2} \\ z_2 &= 1/y_2 = (z_{20}z_{2s})^{1/2} = 1/(y_{20}y_{2s})^{1/2} \end{aligned} \tag{7.23}$$

Now, let us introduce a dimensionless factor δ which is the ratio of the open- and short-circuit impedances

$$\delta^2 = z_{10}/z_{1s} = z_{20}/z_{2s} \tag{7.24}$$

Then

$$\delta = \pm(z_{10}y_{1s})^{1/2} = \pm(z_{20}y_{2s})^{1/2} \tag{7.25}$$

and also

$$\begin{aligned} z_{11}{}^b &= z_{10} = \delta z_1 \\ z_{22}{}^b &= z_{20} = \delta z_2 \\ y_{11}{}^a &= y_{1s} = \delta y_1 \\ y_{22}{}^a &= y_{2s} = \delta y_2 \end{aligned} \tag{7.26}$$

which can be used to define the sign to be associated with δ. (In the sequel, the positive sign will generally be employed.)

With the parameters z_1, z_2, and δ, the mutual immitances $z_{12}{}^b$ and $y_{12}{}^a$ can be expressed as

$$\begin{aligned} z_{12}{}^b &= (z_1 z_2)^{1/2}(\delta^2 - 1)^{1/2} \\ y_{12}{}^a &= (y_1 y_2)^{1/2}(\delta^2 - 1)^{1/2} \end{aligned} \tag{7.27}$$

Thus, it is possible to specify all parameters of the two-terminal-pair in terms of the three quantities z_1, z_2, and δ.

With the admitted use of hindsight, let us now define a new function ϕ dependent upon δ via the hyperbolic functions as

$$\delta = \coth \phi = \frac{\cosh \phi}{\sinh \phi} = \frac{e^{\phi} + e^{-\phi}}{e^{\phi} - e^{-\phi}} \tag{7.28}$$

from which

$$e^{2\phi} = e^{2(\alpha + j\beta)} = \frac{\delta + 1}{\delta - 1} \tag{7.29}$$

where ϕ, being a function of $j\omega$ in the steady state, has real and imaginary parts. Other useful relations are

$$\begin{aligned} \delta^2 - 1 &= \coth^2 \phi - 1 = \operatorname{csch}^2 \phi = 1/\sinh^2 \phi \\ \cosh \phi &= \frac{\delta}{(\delta^2 - 1)^{1/2}} \end{aligned} \tag{7.30}$$

and

$$\begin{aligned} z_{10} &= z_1 \coth \phi & z_{20} &= z_2 \coth \phi \\ y_{1s} &= y_1 \coth \phi & y_{2s} &= y_2 \coth \phi \\ z_{12}{}^b &= (z_1 z_2)^{1/2} \operatorname{csch} \phi & y_{12}{}^a &= (y_1 y_2)^{1/2} \operatorname{csch} \phi \\ A &= (z_1/z_2)^{1/2} \cosh \phi & B &= (z_1 z_2)^{1/2} \sinh \phi \\ C &= (y_1 y_2)^{1/2} \sinh \phi & D &= (z_2/z_1)^{1/2} \cosh \phi \end{aligned} \tag{7.31}$$

The various equations relating e_1, e_2, i_1, and i_2 can easily be expressed in terms of the image impedances z_1 and z_2 and the function ϕ. A rather noticeable symmetry exists in these expressions. Of most immediate interest are the A, B, C, D relations. We get

$$\begin{aligned} e_1 &= Ae_2 - Bi_2 = [(z_1/z_2)^{1/2} \cosh \phi]e_2 - [(z_1 z_2)^{1/2} \sinh \phi]i_2 \\ i_1 &= Ce_2 - Di_2 = [(y_1 y_2)^{1/2} \sinh \phi]e_2 - [(z_2/z_1)^{1/2} \cosh \phi]i_2 \end{aligned} \tag{7.32}$$

Now let us assume that there exists a load impedance Z on the network. Then, the voltage e_2 is given by

$$e_2 = -i_2 Z = -i_2/Y \tag{7.33}$$

where the negative sign accounts for the assumed current direction. Using eq. 7.33 to eliminate i_2 in the first of eqs. 7.32 and e_2 in the second, there is obtained

$$\begin{aligned} e_1/e_2 &= (z_1/z_2)^{1/2}(\cosh\phi + z_2 Y \sinh\phi) \\ i_1/i_2 &= -(z_2/z_1)^{1/2}(\cosh\phi + y_2 Z \sinh\phi) \end{aligned} \tag{7.34}$$

Let us take the obvious suggestion evident in eqs. 7.34 and terminate the network in its image impedance z_2. Then, we take $Z = z_2$ and eqs. 7.34 assume the rather remarkable form

$$e_1/e_2 = (z_1/z_2)^{1/2}e^{\phi} \qquad i_1/i_2 = -(z_2/z_1)^{1/2}e^{\phi} \tag{7.35}$$

For a physically symmetric network, $z_1 = z_2$ so that the voltage ratio becomes simply e^{ϕ} while the current ratio becomes simply $-e^{\phi}$.

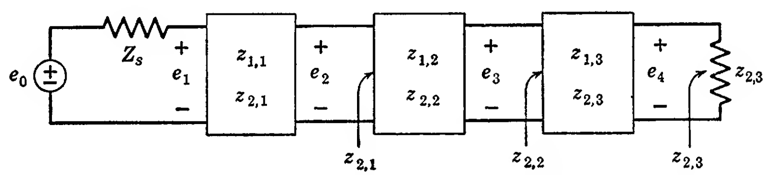


Fig. 7.7. The first requirement for an image-matched filter chain.

Let us now consider a chain of networks, each of which affords an image termination on the preceding network as shown in Fig. 7.7. Because each network is properly terminated in its image impedance

$$e_1/e_2 = (z_{1,1}/z_{2,1})^{1/2}e^{\phi} \tag{7.36}$$

with similar equations for the ratios e_3/e_4 and e_2/e_3. Putting these three functions together

$$\frac{e_1}{e_4} = \left(\frac{z_{1,1}z_{1,2}z_{1,3}}{z_{2,1}z_{2,2}z_{2,3}}\right)^{1/2} e^{(\phi_1+\phi_2+\phi_3)} \tag{7.37}$$

In other words, *the consistent use of image terminations enables us to find the reverse function of several networks in cascade as the product of the reverse functions of the individual networks.*

Let us divide one of eqs. 7.35 by the other. The result is

$$\frac{e_1 i_2}{e_2 i_1} = -\frac{z_1}{z_2} \tag{7.38}$$

But since e_2 and i_2 are related according to Ohm's law as in eq. 7.33, eq. 7.38 becomes

$$e_1/i_1 = z_1 \tag{7.39}$$

In other words, *when a network is terminated in its image impedance z_2, the input impedance to the network is the image impedance z_1.* This means that $z_{2,2} = z_{1,3}$, $z_{2,1} = z_{1,2}$, and so forth, in Fig. 7.7.

In order that the voltage e_1 of Fig. 7.7 be reasonably related to the source voltage e_0, it is necessary that Z_s be related to the input im-

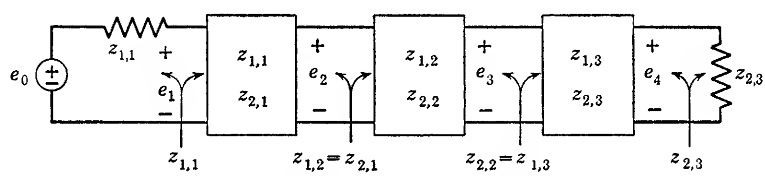


Fig. 7.8. The general image-matched system.

pedance of the first network. Since the first network is terminated in $z_{2,1}$, its input impedance is $z_{1,1}$. The most obvious choice is to make $Z_s = z_{1,1}$ so that $e_1 = e_0/2$. If $Z_s = z_{1,1}$, the impedance looking into the first network from the right will be $z_{2,1}$ because a signal travelling left will see an image-impedance match much the same as a signal travelling right. Thus, we are finally led to the network of Fig. 7.8. The striking characteristic of this network is that at the junction between each individual network the *same* impedance is seen looking either way.

In general, an image match is not optimum for power transfer. However, when the image impedances are purely resistive, an image match is also a conjugate match.

We may find the reverse function for the complete network of Fig. 7.8 as

$$\frac{e_0}{e_4} = 2\left(\frac{z_{1,1}z_{1,2}z_{1,3}}{z_{2,1}z_{2,2}z_{2,3}}\right)^{1/2} e^{(\phi_1+\phi_2+\phi_3)} \tag{7.40}$$

If each individual network is physically symmetric, the network of Fig. 7.8 simplifies to that of Fig. 7.9 and the over-all reverse function

becomes

$$\frac{e_0}{e_4} = 2e^{(\phi_1+\phi_2+\phi_3)} = 2[e^{(\alpha_1+\alpha_2+\alpha_3)}][e^{j(\beta_1+\beta_2+\beta_3)}] \tag{7.41}$$

Even though all the image impedances for the symmetric system are equal, this does not require that the reverse functions of the individual networks be the same. Even in the symmetric system (the one of most practical importance), almost complete variety in the design of the individual networks is possible.

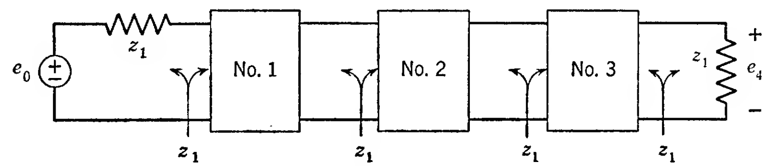


Fig. 7.9. The image-matched system with symmetric filter sections.

7.4 Insertion ratios

In discussions of bilateral filter chains, the concepts of voltage, current, and power insertion ratios play an important part. As implied by the name, an insertion ratio gives the behavior of a network placed between source and load impedances in terms of the behavior of the same source and load impedances without an intervening network. The nomenclature is described by Fig. 7.10. In *a*, no network exists, whereas in *b* the system with an intervening network is shown. The insertion ratio is defined as the reverse function of the system with the filter divided by the reverse function without the filter. Accordingly,

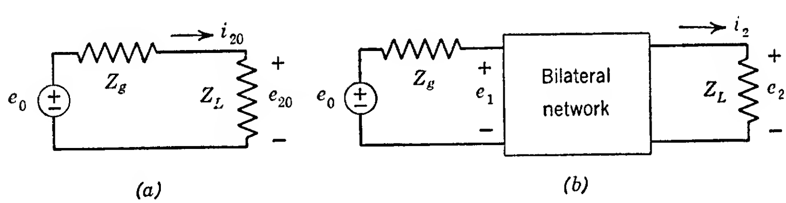


Fig. 7.10. Notation for defining insertion ratios.

$$\begin{aligned} \text{Insertion voltage ratio} &= e_{20}/e_2 \\ \text{Insertion current ratio} &= i_{20}/i_2 \\ \text{Insertion power ratio} &= P_{20}/P_2 \end{aligned} \tag{7.42}$$

The power in the load is

$$P_{20} = \frac{|e_{20}|^2}{\text{Re } Z_L} = |i_{20}|^2 \text{ Re } Z_L$$
$$P_2 = \frac{|e_2|^2}{\text{Re } Z_L} = |i_2|^2 \text{ Re } Z_L \tag{7.43}$$

where root-mean-square values have been assumed. Thus

$$P_{20}/P_2 = |e_{20}/e_2|^2 = |i_{20}/i_2|^2 \tag{7.44}$$

which is just the square of the magnitude of the voltage or current insertion ratio. The insertion power ratio (a loss) can be expressed in terms of decibels according to

$$\begin{aligned} \text{Decibels} &= 10 \log_{10} (P_{20}/P_2) \\ &= 20 \log_{10}|e_{20}/e_2| = 20 \log_{10}|i_{20}/i_2| \end{aligned} \tag{7.45}$$

Let us define the insertion voltage ratio in terms of an exponential. Then

$$e^\theta = \frac{e_{20}}{e_2} = \frac{Z_L}{Z_L + Z_g}\left(\frac{e_0}{e_2}\right) \tag{7.46}$$

which, for the perfect image-matched system, is simply related to the reverse function e^ϕ as

$$e^\theta = \frac{2(z_1 z_2)^{1/2}}{z_1 + z_2} e^\phi \tag{7.47}$$

where z_1 and z_2 are the image-impedance source and load respectively. For the symmetric image-matched system where $z_1 = z_2$, we have very simply $e^\theta = e^\phi$; that is, *the reverse function is also the insertion voltage ratio for symmetric image-matched networks.*

The function θ can be expressed in terms of real and imaginary parts $\alpha + j\beta$. Therefore

$$e^{2\alpha} = |e^\theta|^2 = \frac{P_{20}}{P_2} = \left|\frac{Z_L}{Z_L + Z_g}\left(\frac{e_0}{e_2}\right)\right|^2 \tag{7.48}$$

The power P_2 can never be larger than P_{20} when $Z_L = Z_g{}^*$. Thus

$$(P_2)_{\max} = \frac{|e_0/2|^2}{\text{Re } Z_g} \tag{7.49}$$

which permits bounds to be placed on the insertion power ratio according to

$$\frac{P_{20}}{P_2} \gtrless \frac{P_{20}}{(P_2)_{\max}} = \frac{4\,\mathrm{Re}\,Z_g}{\mathrm{Re}\,Z_L}\left|\frac{Z_L}{Z_L + Z_g}\right|^2 \tag{7.50}$$

If the source and load impedances have resistive parts R_g and R_L respectively

$$\frac{P_{20}}{P_2} \gtrless \frac{4R_gR_L}{(R_g + R_L)^2} \tag{7.51}$$

In a low-pass filter at $\omega = 0$, all the capacitors in the network are open circuits and all the inductors are short circuits. If the network is purely reactive (except for the source and load resistors), $P_{20} = P_2$ at $\omega = 0$. It is impossible to obtain a match between unequal resistors in a low-pass lossless filter at and near $\omega = 0$, nor can practical transformers help because practical transformers do not pass very low frequencies. A transformer or some arrangement of inductors and capacitors for matching different resistances works only at finite frequencies. Accordingly, the design of conventional low-pass lossless filters reasonably assumes $R_g = R_L$ so that $P_{20}/P_2 \gtrless 1$.

For band-pass or band-elimination filters, it is possible to assume $R_g = R_L$ and then utilize determinant manipulation if the source and load resistances are different. Note, however, that an approximate image-impedance match (and hence an approximate conjugate match when source and load are resistances) is required. Therefore, the benefits of intentional mismatch (which includes the ideal current source) cannot be realized with image-matched systems.

More often than not, low-pass filters need not pass very low frequencies (below perhaps 60 cycles in speech systems). Then, practical transformers may be employed in lieu of ideal transformers in order to match resistors differing by any reasonable amount. Of course, the match at and near $\omega = 0$ will be poor, although above a few cycles per second it may be quite good.

7.5 The effect of mismatch

Obtaining a perfect image-impedance match is often not practical as when the image impedance is a fairly complicated function of frequency. Ideally, resistors should be employed as source and load impedances. One type of image-matched filter ("constant-resistance" network) has an image impedance that is resistive and constant for all frequencies. Simple source and load resistors with such networks result in perfect image terminations. However, constant-resistance networks are often

not practical or desirable because of their limitations. In fact, if a capacitance occurs in shunt with the input or the output of a network, the network can never be made constant resistance. (In addition, constant-resistance networks require more circuit elements than other types for a given transfer function.) Then, image-matched systems in which the image impedances are only approximately constant resistances must be employed and the terminations must be resistances that afford a match only *on the average* over the frequency band of interest. Obvi-

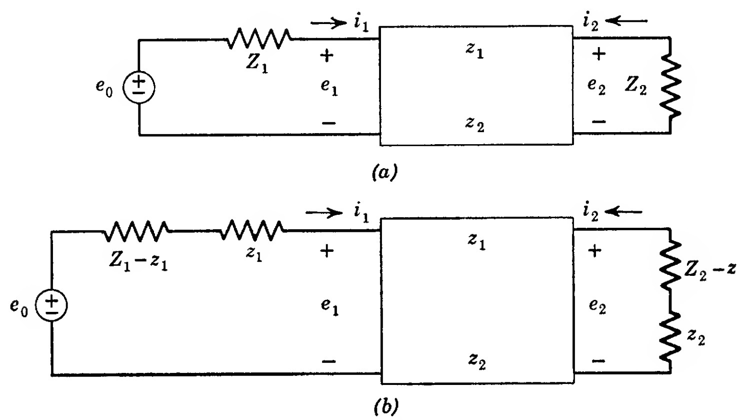


Fig. 7.11. Network development for determining mismatch effects.

ously, it is important to be able to calculate the deviation of the actual behavior of the network from its ideal behavior calculated assuming perfect image terminations.

Consider the image-derived network of Fig. 7.11a in which the source and load impedances are not equal to the image impedances z_1 and z_2 respectively. Let us rearrange the terminating impedances as shown in Fig. 7.11b; no change in the circuit has actually been made. The current flowing in the impedance $Z_1 - z_1$ produces a voltage $i_1(Z_1 - z_1)$, and that flowing in $Z_2 - z_2$ yields a voltage $i_2(Z_2 - z_2)$. We can re-

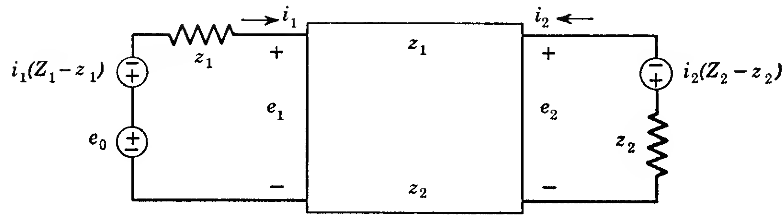


Fig. 7.12. Final network for determining mismatch effects.

place the relevant impedances with equivalent voltage generators as shown in Fig. 7.12. The value of this artifice is that now three voltages operate in a perfect image-matched system and the contribution to the over-all behavior given by each becomes rather simple to obtain; superposition can be employed to find the actual response of the mismatched system.

The current i_2 is made up of two components. The first i_{2a} results from the two voltage generators at end 1 of the network. From the reverse function

$$i_1/i_2 = -(z_2/z_1)^{1/2}e^{\phi} \tag{7.52}$$

we obtain the first component of current as

$$i_{2a} = -i_{1a}(z_1/z_2)^{1/2}e^{-\phi} \tag{7.53}$$

where i_{1a} is the component of current i_1 resulting only from the two voltage generators at end 1.

Since

$$i_{1a} = \frac{e_0 - i_1(Z_1 - z_1)}{2z_1} \tag{7.54}$$

eq. 7.53 becomes

$$i_{2a} = -\frac{e_0 - i_1(Z_1 - z_1)}{2(z_1z_2)^{1/2}}e^{-\phi} \tag{7.55}$$

The second component of i_2 comes from the voltage generator at end 2 of the network

$$i_{2b} = -\frac{i_2(Z_2 - z_2)}{2z_2} \tag{7.56}$$

Thus, the net current i_2 is

$$i_2 = i_{2a} + i_{2b} = -\frac{[e_0 - i_1(Z_1 - z_1)]e^{-\phi}}{2(z_1z_2)^{1/2}} - \frac{i_2(Z_2 - z_2)}{2z_2} \tag{7.57}$$

The first component of i_1 is given by eq. 7.54. The second component of i_1 is given by eq. 7.55 after exchanging subscripts (by symmetry) and setting $e_0 = 0$. Thus

$$i_1 = i_{1a} + i_{1b} = \frac{e_0 - i_1(Z_1 - z_1)}{2z_1} + \frac{i_2(Z_2 - z_2)e^{-\phi}}{2(z_1z_2)^{1/2}} \tag{7.58}$$

Equations 7.57 and 7.58 are two equations in the two unknowns i_1 and i_2; hence, it is possible to solve for the two currents. In particular, we can solve for i_2. If the network is removed and a direct connection

substituted between source and load, i_2 becomes

$$i_{20} = -e_0/(Z_1 + Z_2) \tag{7.59}$$

which we can use to form the insertion current ratio as

$$\frac{i_{20}}{i_2} = \left(\frac{2(Z_1Z_2)^{1/2}}{Z_1 + Z_2}\right)\left(\frac{\sigma}{F_{1i}F_{2i}}\right)e^{\phi} \tag{7.60}$$

where F_{1i} measures the match between Z_1 and z_1, F_{2i} measures the match between Z_2 and z_2, and σ contains mixed terms. The first term of eq. 7.60 in parenthesis is unity if the load and source impedances are equal. The parameters of importance are

$$\begin{aligned} F_{1i} &= \frac{2(Z_1z_1)^{1/2}}{Z_1 + z_1} \\ F_{2i} &= \frac{2(Z_2z_2)^{1/2}}{Z_2 + z_2} \\ \sigma &= 1 - \left(\frac{Z_1 - z_1}{Z_1 + z_1}\right)\left(\frac{Z_2 - z_2}{Z_2 + z_2}\right)e^{-2\phi} \end{aligned} \tag{7.61}$$

If $Z_1 = z_1$ or if $Z_2 = z_2$ or if $e^{-2\alpha}$ is large (large attenuation), the mixed term σ is unity. For the matched case where $Z_1 = z_1$ and $Z_2 = z_2$, eq. 7.60 reduces to the ideal insertion ratio of eq. 7.47.

In lossless filters built in the form of a ladder network, the mixed term may oscillate in magnitude somewhat over certain frequency regions. This can result in a rippled response. However, when many filter sections make up a single composite filter, incidental dissipation may be sufficient to make σ little different from unity.

7.6 Image parameters of specific networks

It is our purpose here to set down the image parameters of the most important two-terminal-pair networks. In the interests of economy and practicability, we shall restrict our attention to symmetric networks except for the L network which we consider first. This network ter-

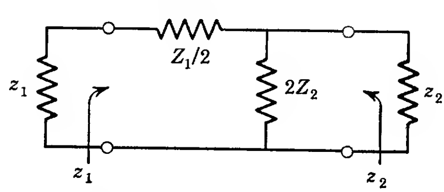


Fig. 7.13. The L network.

minated in its image impedances is shown in Fig. 7.13. Note the factors $\frac{1}{2}$ and 2 used in defining the impedances. The reason for doing this will become apparent later. We find

$$\begin{aligned} z_1 &= Z_T = (z_{10}z_{1s})^{1/2} = (Z_1Z_2)^{1/2}(1 + Z_1/4Z_2)^{1/2} \\ z_2 &= Z_\pi = (z_{20}z_{2s})^{1/2} = (Z_1Z_2)^{1/2}/(1 + Z_1/4Z_2)^{1/2} \\ \coth\phi &= \delta = (z_{10}y_{1s})^{1/2} = (z_{20}y_{2s})^{1/2} = (1 + 4Z_2/Z_1)^{1/2} \\ \cosh\phi &= \delta/(\delta^2 - 1)^{1/2} = (1 + Z_1/4Z_2)^{1/2} \\ (e^{2\phi})_L &= \frac{\delta + 1}{\delta - 1} = \frac{(1 + 4Z_2/Z_1)^{1/2} + 1}{(1 + 4Z_2/Z_1)^{1/2} - 1} \end{aligned} \tag{7.62}$$

The reasons for the new definitions for z_1 and z_2 (Z_T and Z_π) will also become apparent shortly.

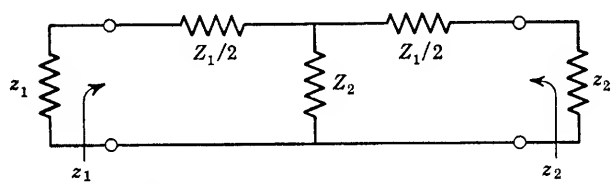


Fig. 7.14. The symmetric T network.

Now consider the symmetric T network of Fig. 7.14. Clearly, $z_1 = z_2$. We find

$$\begin{aligned} z_1 = z_2 &= Z_T = (Z_1Z_2)^{1/2}(1 + Z_1/4Z_2)^{1/2} \\ \cosh\phi &= 1 + Z_1/2Z_2 \end{aligned} \tag{7.63}$$

We observe that the image impedance of the T network is the same as one of the impedances associated with the L network. In fact, we can put two L networks back to back to get the T network. The image transfer constant ϕ for the L network is just half that of the T network so that

$$(e^{2\phi})_T = (e^{2\phi})_L{}^2 = \frac{1 + 2Z_2/Z_1 + (1 + 4Z_2/Z_1)^{1/2}}{1 + 2Z_2/Z_1 - (1 + 4Z_2/Z_1)^{1/2}} \tag{7.64}$$

Now consider the symmetric pi network of Fig. 7.15. With the aid of the general equations, we get

$$\begin{aligned} z_1 = z_2 &= Z_\pi = (Z_1Z_2)^{1/2}/(1 + Z_1/4Z_2)^{1/2} \\ \cosh\phi &= 1 + Z_1/2Z_2 \\ (e^{2\phi})_\pi &= (e^{2\phi})_T \end{aligned} \tag{7.65}$$

As before, we see that two L networks placed together make up a pi network. The factors $\frac{1}{2}$ and 2 were introduced so that this could be done directly.

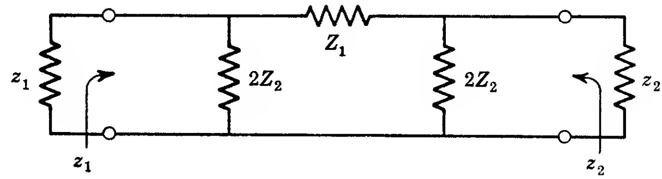


Fig. 7.15. The symmetric pi network.

Because Z_T is seen looking into one end of the L network and Z_π looking into the other, the L network can match T and pi sections when some complicated filter chain is built up from L, T, and pi networks in cascade. An example of a relatively complicated image-matched chain is shown in Fig. 7.16 in which the matching criterion is that all Z_T are the same and all Z_π are the same. If the networks have image quantities ϕ_1, ϕ_2, and so forth, then when everything is properly matched $e_1 = e_0/2$ and the over-all reverse function is

$$\frac{e_0}{e_n} = 2e^{(\phi_1+\phi_2+\phi_3+\cdots)} \tag{7.66}$$

Finally, consider the symmetric lattice network of Fig. 7.17*a*. This network is only indirectly related to the T and pi, although equivalences do exist which will be brought out in the following section. The normal way of drawing the symmetric lattice is shown in Fig. 7.17*b*. Again with the aid of general equations containing short- and open-circuit immittances, we get

$$\begin{aligned}
z_1 &= z_2 = (Z_aZ_b)^{1/2} \\
\cosh\phi &= \frac{1 + Z_a/Z_b}{1 - Z_a/Z_b} \\
\coth\phi &= \delta = \frac{(Z_a/Z_b)^{1/2} + (Z_b/Z_a)^{1/2}}{2} \\
(e^\phi)_{\text{Lat}} &= \frac{1 + (Z_a/Z_b)^{1/2}}{1 - (Z_a/Z_b)^{1/2}}
\end{aligned} \tag{7.67}$$

The simplicity and symmetry of the lattice equations make the lattice the easiest network with which to deal. This simplicity is indeed fortunate because the lattice is also the most general network. In fact, *if*

a transfer function can be realized with any arbitrary symmetric network, it can be realized with a symmetric lattice network. It will be noted that z_1 is dependent upon the product of Z_a and Z_b, whereas e^{ϕ} is dependent upon the ratio of Z_a and Z_b. This means that the image impedance

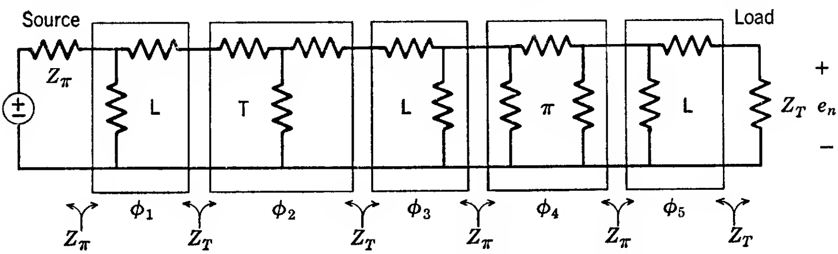


Fig. 7.16. A typical ladder filter with L, T, and pi sections.

and the reverse function of the lattice can be specified independently, which is basically the reason why the lattice is so general a network.

Although lattice networks can be used in a physical system, it is often more practical to employ unbalanced networks such as the T, pi,

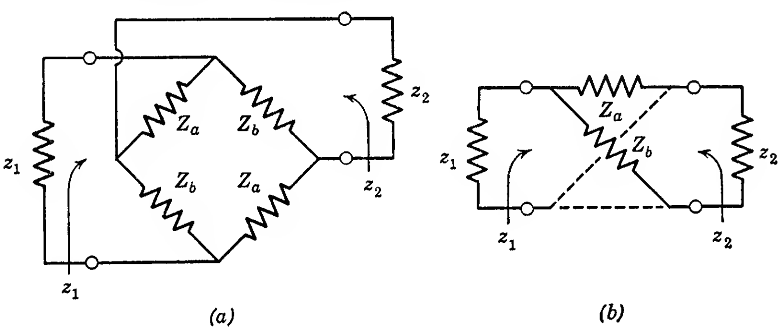


Fig. 7.17. The symmetric lattice network.

or L. In many cases of practical importance, lattice networks can be assumed in the design of a filter and then the individual lattices can be "decomposed" into unbalanced equivalents. The lattice is not always desirable for two reasons. First, it requires more circuit elements than are necessary (thus, it is not an economical network); and second, much of its behavior is dependent upon a bridge type of balance which requires rather accurately specified and adjusted circuit element values.

7.7 Relationships between the lattice and other networks

Any arbitrary but symmetric two-terminal-pair network has an equivalent lattice network in so far as the two pairs of terminals are

concerned. This statement is called Bartlett's bisection theorem and shows just how general the lattice network really is. Because of its importance, we shall prove part of the theorem.

Let some symmetric two-terminal-pair network be halved so that each half is the same and so that the connecting wires between the two halves are available. This is described by Fig. 7.18. Two cases present

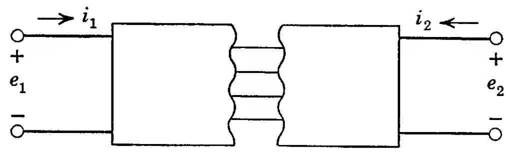


Fig. 7.18. The halved symmetric two-terminal-pair network.

themselves depending upon whether the wires connecting the two halves cross or do not cross. We shall prove the latter case and merely state the results for crossing wires.

Let us prove the bisection theorem by studying the equivalent T network of Fig. 7.2 (which, although a valid equivalent for any network, will not necessarily be physically realizable). Drawing this network apart (assuming it to be symmetric), we get the circuit of Fig. 7.19.

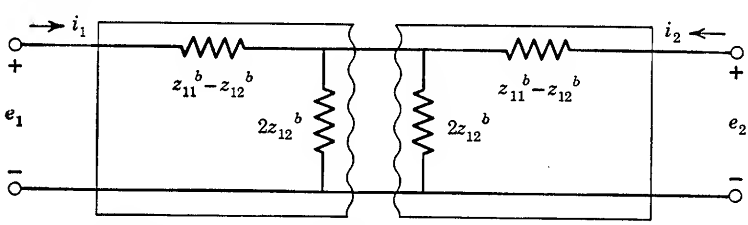


Fig. 7.19. The halved general symmetric T equivalent.

The input impedance to this network when the interconnecting wires are open-circuited will be called Z_{och} (standing for open-circuited half), and when the interconnecting wires are short-circuited the input impedance is Z_{sch} (standing for short-circuited half). Thus

$$\begin{aligned} Z_{\text{och}} &= z_{11}^b + z_{12}^b \\ Z_{\text{sch}} &= z_{11}^b - z_{12}^b \end{aligned} \tag{7.68}$$

But these impedances are easily verified to be the impedances of the two arms of a symmetric lattice. Therefore

$$\begin{aligned} Z_a \text{ (or } Z_b) &= Z_{\text{sch}} \\ Z_b \text{ (or } Z_a) &= Z_{\text{och}} \end{aligned} \tag{7.69}$$

which are *always* realizable if the original symmetric network physically exists. (Whether we take Z_{sch} to represent Z_a or Z_b depends on definition and whether or not one pair of terminals to the network is reversed.)

Equations 7.69 can be proved by substituting for $z_{11}{}^a$ and $z_{12}{}^b$ in eqs. 7.68. For the symmetric lattice, $z_{11}{}^b = z_{10} = z_{20} = (Z_a + Z_b)/2$.

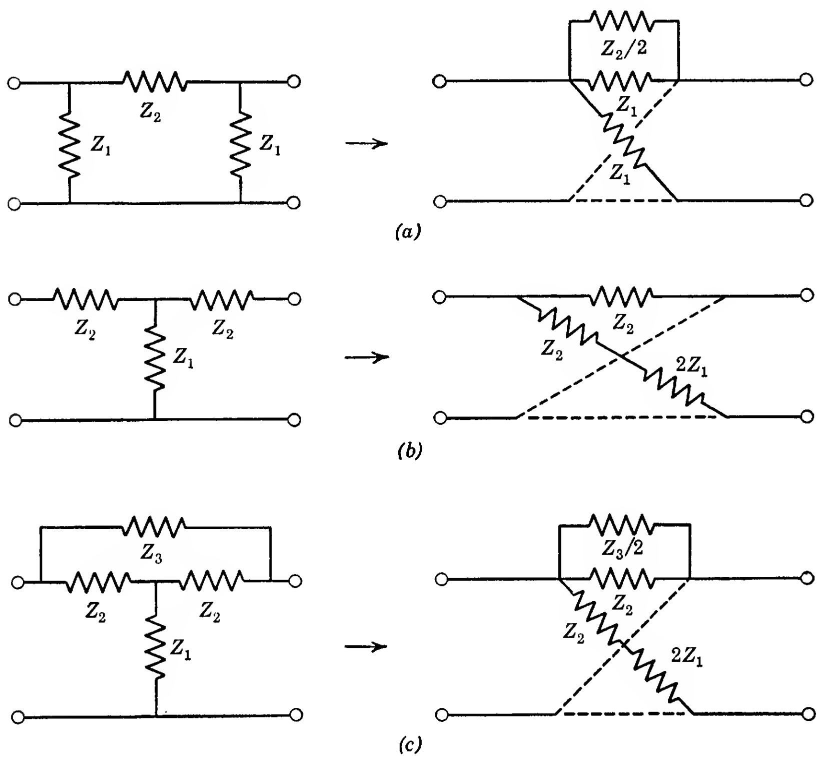


Fig. 7.20. Lattice equivalents.

Also, $y_{11}{}^a = y_{1s} = y_{2s} = (Y_a + Y_b)/2$. Substituting these values into eqs. 7.68 (making use of eqs. 7.9) results in eqs. 7.69.

Lattice equivalents to symmetric T and pi networks and to the bridged-T network are shown in Fig. 7.20.

When some of the interconnecting wires of the drawn-apart symmetric network cross, the equivalent symmetric lattice arms can also be found *providing* that the network also has symmetry about a horizontal line drawn through the center of the filter. (The case of crossed wires enables us, for example, to find the lattice equivalent to a lattice or cascade of lattice networks.) Then, Z_a is the input impedance to half the net-

work with the interconnecting wires that do not cross short-circuited together as before but with the wires that do cross left open-circuited. Z_b is the input impedance with the noncrossing wires open as before but with the crossing wires short-circuited together.

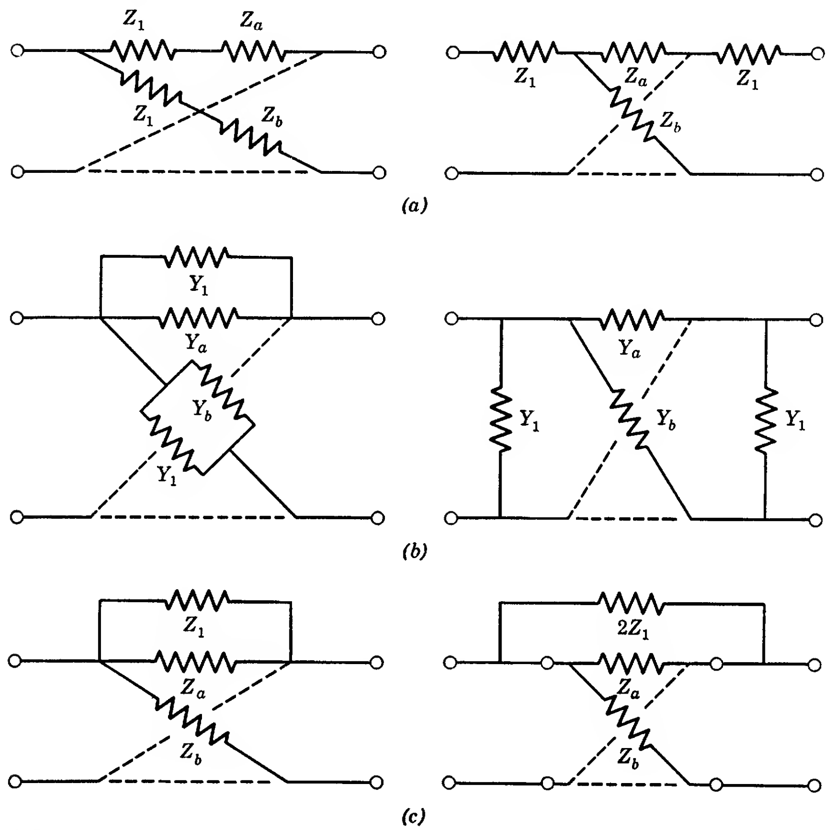


Fig. 7.21. Decomposing techniques. (*a*) Common series impedance. (*b*) Common shunt admittance. (*c*) Bridged impedance.

Although the lattice equivalent of a symmetric ladder can always be found as a physically meaningful network, the converse is not always true. However, many lattices do have practical ladder equivalents; consequently, it is of interest to see what can be done in the way of "decomposing" a lattice into a practical ladder. Various decomposing tricks are shown in Fig. 7.21. The proof of these is left to the reader.

As a practical example, consider Fig. 7.22 which shows the structure of the two arms of a lattice. First we note that a shunt C_1 is a common first element in Z_a and Z_b. Therefore we can "pull out" this capacitor.

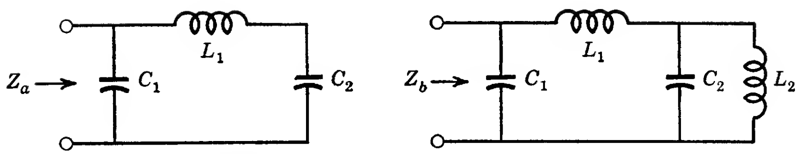


Fig. 7.22. Lattice arms for the example.

After doing this, we observe that the remaining lattice arms have a common series inductance which can be pulled out. The result of these two steps is the network of Fig. 7.23*a*. In one arm of the lattice remaining in Fig. 7.23*a*, a capacitor C_2 is in parallel with an infinite im-

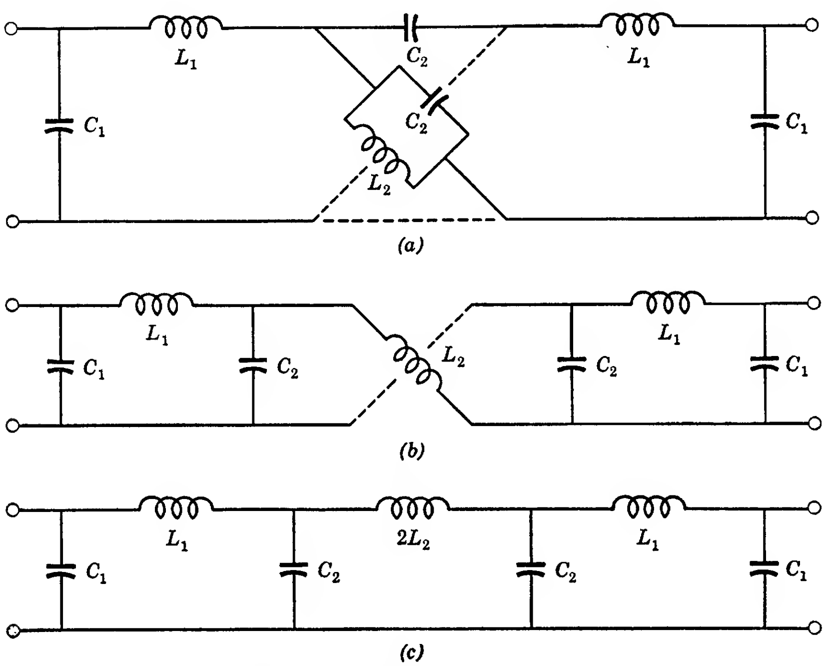


Fig. 7.23. Example of decomposition.

pedance (an open circuit), whereas in the other arm the same element is in shunt with L_2. Therefore we again have a common shunt element and the network further decomposes to that of Fig. 7.23*b*. We are now left with a lattice consisting simply of crossed inductors. Thus the network of Fig. 7.23*b* becomes the practical ladder of Fig. 7.23*c*, the only difference being a reversal of polarity between input and output as compared to the original lattice because the two inductances are uncrossed in the last step of the decomposition procedure.

Problems

1. Derive eqs. 7.6.
2. Derive eqs. 7.7.
3. Derive eqs. 7.9.
4. Derive eq. 7.10.
5. Derive eqs. 7.11.
6. Derive eqs. 7.12.
7. Derive eqs. 7.13.
8. Show that the general T and pi networks of Fig. 7.2 are valid equivalents.
9. Derive eqs. 7.15.
10. Derive eqs. 7.17.
11. Derive eq. 7.19.
12. Derive eqs. 7.22.
13. Derive eqs. 7.26.
14. Derive eqs. 7.27.
15. Derive eq. 7.29.
16. Derive eqs. 7.30.
17. Derive eqs. 7.31.
18. Derive eqs. 7.32.
19. Derive eq. 7.47.
20. Derive eq. 7.60.
21. Derive eqs. 7.61.
22. Show how two L networks make up a T network.

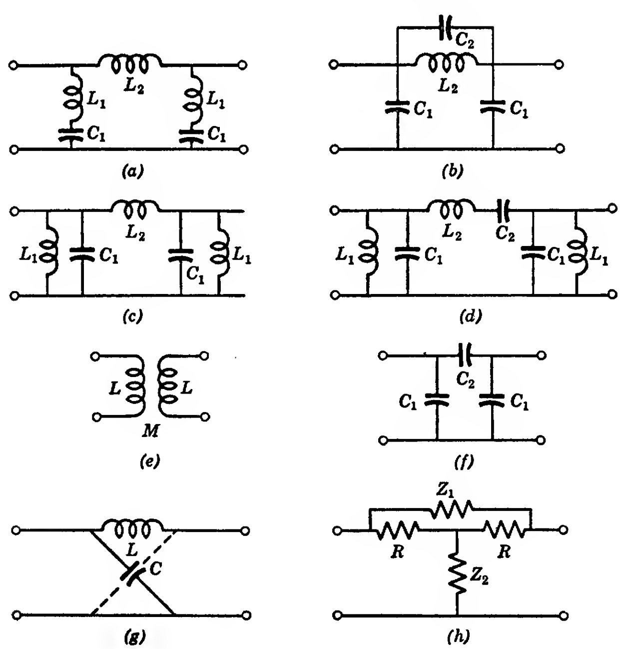


Fig. P.29.

23. Show how two L networks make up a pi network.
24. Derive eqs. 7.62.
25. Derive eqs. 7.63.
26. Derive eqs. 7.65.
27. Derive eq. 7.67.
28. Show that eqs. 7.69 are true.
29. Find the lattice equivalents of the networks of Fig. P.29.
30. Prove that the networks of Fig. 7.21*a* are equivalent.
31. Prove that the networks of Fig. 7.21*b* are equivalent.
32. Prove that the networks of Fig. 7.21*c* are equivalent.

8

Conventional Filters

In this chapter, we shall use the image-impedance theory and other data from the preceding chapter to arrive at specific design procedures for a multitude of filters containing T, pi, L, and lattice sections. Ladder filters, being of the greatest practical importance, will be treated in the most detail.

The methods described here admit of sophistication which we simply do not have the time to pursue. In fact, Chebyshev and other types of functions can be developed as an extension of conventional filter theory. For discussions of such topics, the reader is referred to other material.

8.1 Transmission and attenuation in T and pi sections

The conditions for stop and pass bands in filters composed of T, pi, and L ladder sections will now be derived. We shall restrict ourselves to symmetric image-matched filters composed of lossless circuit elements; that is, it will be assumed that, except for source and load impedances, only coils and capacitors and possibly transformers make up the filter.

For either pi or T sections and with the nomenclature of Sec. 7.6

$$\begin{aligned} \cosh \phi &= 1 + Z_1/2Z_2 \\ Z_T &= (Z_1 Z_2)^{1/2}(1 + Z_1/4Z_2)^{1/2} \\ Z_\pi &= (Z_1 Z_2)^{1/2}/(1 + Z_1/4Z_2)^{1/2} \end{aligned} \tag{8.1}$$

Because Z_1 and Z_2 have been assumed to be purely reactive

$$\begin{aligned} &Z_1/2Z_2 \text{ is purely real} \\ &Z_1/4Z_2 \text{ is purely real} \\ &(Z_1 Z_2)^{1/2} \text{ is purely real or purely imaginary} \end{aligned} \tag{8.2}$$

The hyperbolic cosine of eqs. 8.1 can be expanded as

$$\begin{aligned}\cosh \phi = \cosh (\alpha + j\beta) &= \cosh \alpha \cos \beta + j \sinh \alpha \sin \beta \\ &= 1 + Z_1/2Z_2\end{aligned} \tag{8.3}$$

Because $Z_1/2Z_2$ is *always* real

$$\sinh \alpha \sin \beta = 0 \tag{8.4}$$

which can be satisfied in one of two ways

$$\alpha = 0 \quad \text{or} \quad \beta = 0, \quad \pm\pi \tag{8.5}$$

The quantity α is the attenuation constant. Therefore $\alpha = 0$ corresponds to unimpeded transmission of signals through the network. In other words, $\alpha = 0$ *defines a pass band.* With $\alpha = 0$ in eqs. 8.3, the phase shift in the pass band is given by

$$\cos \beta = 1 + Z_1/2Z_2 \tag{8.6}$$

The cosine of the angle β can certainly not exceed ± 1; should β become complex, α could not then be zero. Therefore, in the pass band

$$\begin{aligned} -1 \leq 1 + Z_1/2Z_2 \leq +1 \\ \text{or} \qquad -1 \leq Z_1/4Z_2 \leq 0 \end{aligned} \tag{8.7}$$

If this permissible range of $Z_1/4Z_2$ is applied to a study of the equations for Z_T and Z_π, we observe that *in a pass band Z_T and Z_π are positive and real* (resistive) *and the attenuation constant α is zero.* Therefore, *in a pass band, Z_1 and Z_2 must be reactances of the opposite types* (that is, if one is capacitive, the other must be inductive).

Now let us study the stop-band conditions of which there are two; one requires $\beta = 0$ and the other $\beta = \pm\pi$. From eqs. 8.3, we find

$$\begin{aligned} \beta = 0: &\quad \cosh \alpha = 1 + Z_1/2Z_2 \\ \beta = \pm\pi: &\quad \cosh \alpha = -(1 + Z_1/2Z_2) \end{aligned} \tag{8.8}$$

and in terms of $Z_1/4Z_2$

$$\begin{aligned} Z_1/4Z_2 \geq 0 &\quad \text{for} \quad \beta = 0 \\ Z_1/4Z_2 \leq -1 &\quad \text{for} \quad \beta = \pm\pi \end{aligned} \tag{8.9}$$

Therefore, in the stop band for $\beta = 0$, Z_1 and Z_2 must be reactances of the same type, whereas in the stop band for $\beta = \pm\pi$, Z_1 and Z_2 will, as in the pass band, be reactances of the opposite types. However, both conditions of eqs. 8.9 lead to the observation that *in a stop band, both Z_T and Z_π are purely reactive.*

The image impedance in a pass band is purely resistive, whereas in a stop band it is purely reactive. If we use simple resistors as source and load impedances, the image match in the pass bands need not be too bad. The match will be poor in the stop bands. However, we are not so concerned about the precise form of signal transmission in stop bands as long as the attenuation is adequate; therefore the problem need not concern us too greatly. The problem of good matches in the pass band is of primary concern.

We may sum up the conditions of $Z_1/4Z_2$ as follows

Stop band	*Pass band*	*Stop band*
$Z_1/4Z_2 < -1$	$-1 < Z_1/4Z_2 < 0$	$0 < Z_1/4Z_2$
$\beta = \pm\pi$	$\alpha = 0$	$\beta = 0$
$\cosh \alpha = -(1 + Z_1/2Z_2)$	$\cos \beta = 1 + Z_1/2Z_2$	$\cosh \alpha = 1 + Z_1/2Z_2$

(8.10)

The attenuation and phase of the image function ϕ are sketched as functions of $Z_1/4Z_2$ (which is always real) in Fig. 8.1. Infinite attenua-

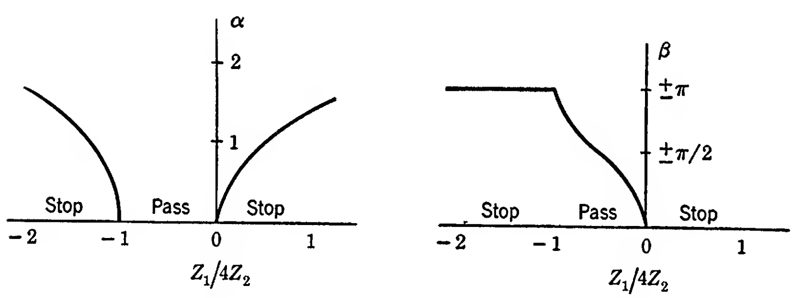


Fig. 8.1. Attenuation and phase of T and pi sections.

tion occurs only if $Z_1/4Z_2$ is either $\pm\infty$ (which occurs when Z_1 or Z_2 is resonant).

The value of $Z_1/4Z_2$ may be plotted as a function of frequency in order to determine the locations in frequency of the various stop and pass bands. This is indicated in Fig. 8.2. By suitable design of reactances Z_1 and Z_2, we may put stop and pass bands any place we please.

The edges of the pass band are defined where $Z_1/4Z_2$ equals 0 and -1. Occasionally, a "characteristic resistance" R_0 for the filter can be defined at the band edge where $Z_1/4Z_2 = 0$. This gives

$$R_0 = (Z_1Z_2)^{1/2} \quad \text{at} \quad Z_1/4Z_2 = 0 \tag{8.11}$$

If the α and β characteristics of Fig. 8.1 refer to a single filter section in a cascade of many properly matched sections, then α and β for the entire filter chain are calculated simply as the sum of the α and β of the individual sections.

Impedances Z_1 and Z_2 of T and pi sections have been assumed to consist only of L's and C's. Therefore these impedances are reactance functions and consequently have p-z only on the $j\omega$ axis of the p plane. The ratio $Z_1/4Z_2$ will thus have p-z only on the $j\omega$ axis. At a pole of Z_1 or a zero of Z_2 (assuming the pole of Z_1 is not also a pole of Z_2), $Z_1/4Z_2$ will become infinite and the attenuation will become infinite.

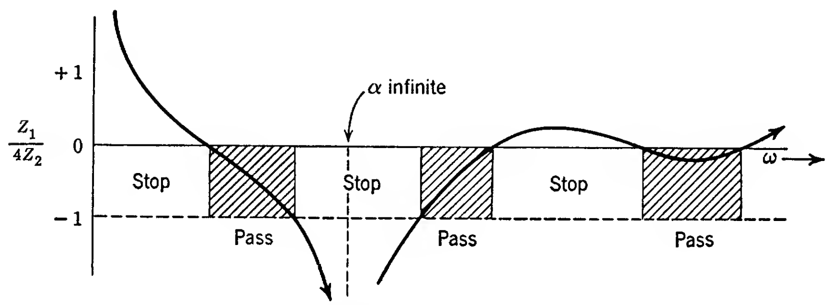


Fig. 8.2. Determining locations of stop and pass bands.

In a physical circuit, this corresponds to a series-resonant circuit in a shunt arm of the T or pi, or to a parallel-resonant circuit in a series arm. At a zero of Z_1 or a pole of Z_2 (if this zero is not also a zero of Z_2), the ratio $Z_1/4Z_2$ is zero, which defines one edge of a pass band. The other edge of the pass band is defined where $Z_1/4Z_2 = -1$, which does not correspond to poles or zeros of the function but rather to the ordinary magnitude of $Z_1/4Z_2$; consequently, this band edge is not so clearly apparent from the p-z plots of Z_1 and Z_2.

An important property is apparent from the p-z concept. That is, if a particular Z_1 and Z_2 comprise a filter, where Z_1 and Z_2 have given p-z, then exchanging the p-z of Z_1 and Z_2 inverts the characteristics of the filter such that the pass bands become stop bands and conversely. A high-pass filter can be derived from a low-pass filter or a band-elimination filter from a band-pass filter in this manner. Exchanging p-z amounts to obtaining reciprocal impedances.

8.2 The simple constant-k low-pass filter

The simplest image-matched T or pi section having reactive elements has a single inductor for Z_1 or Z_2 and a single capacitor for Z_2 or Z_1. Then, the product Z_1Z_2 is independent of frequency and the image-

matched filter is said to be a "constant-k" filter. If $Z_1 = j\omega L$ and $Z_2 = 1/j\omega C$, the filter is of the low-pass type, whereas if $Z_2 = j\omega L$ and $Z_1 = 1/j\omega C$, it is a high-pass filter. The p-z for these two cases are shown in Fig. 8.3. Since the high-pass filter is derivable from the low-

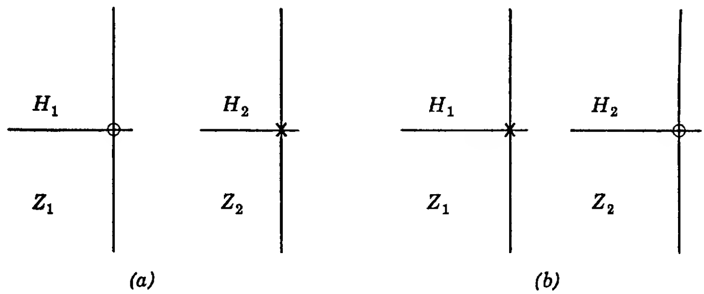


Fig. 8.3. P-z of simple filters. (a) Low pass. (b) High pass.

pass filter (as are also certain types of band-pass and band-elimination filters), we shall speak here only of the low-pass filter.

The ratio of impedances of central importance is

$$\frac{Z_1}{4Z_2} = \frac{j\omega L}{4/(j\omega C)} = -\frac{\omega^2 LC}{4} \tag{8.12}$$

One band edge is defined where this ratio is zero, which occurs at $\omega = 0$ as would be expected. The other band edge is given where the ratio is -1. Letting the corresponding radian frequency be B

$$B = 2/(LC)^{1/2} \tag{8.13}$$

At the band edge defined where the ratio of impedances is zero, we can define a characteristic impedance as

$$R_0 = (Z_1 Z_2)^{1/2} = (L/C)^{1/2} \tag{8.14}$$

A multitude of these simple filters can be designed quite easily. The basic L, T, and pi sections are shown in Fig. 8.4 and various filter chains

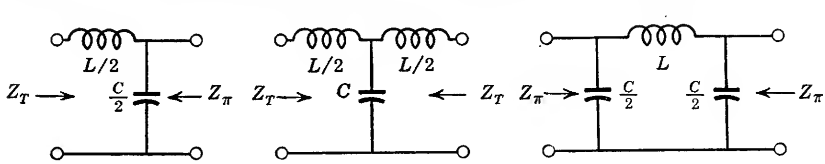


Fig. 8.4. Constant-k low-pass sections.

in Fig. 8.5. The attenuation and phase quantities α and β will be n times that of one T or pi section, where n is the total number of sections (two L sections make up a full section). The bandwidth in all cases is $2/(LC)^{1/2}$ and the characteristic resistance is $(L/C)^{1/2}$. The last two examples of Fig. 8.5 are not symmetric; an extra L section (a "half-section") has been used such that n is the number of full T or pi sections plus one-half. The normalized filter with $R = L = C = 1$ has a bandwidth of 2 radians.

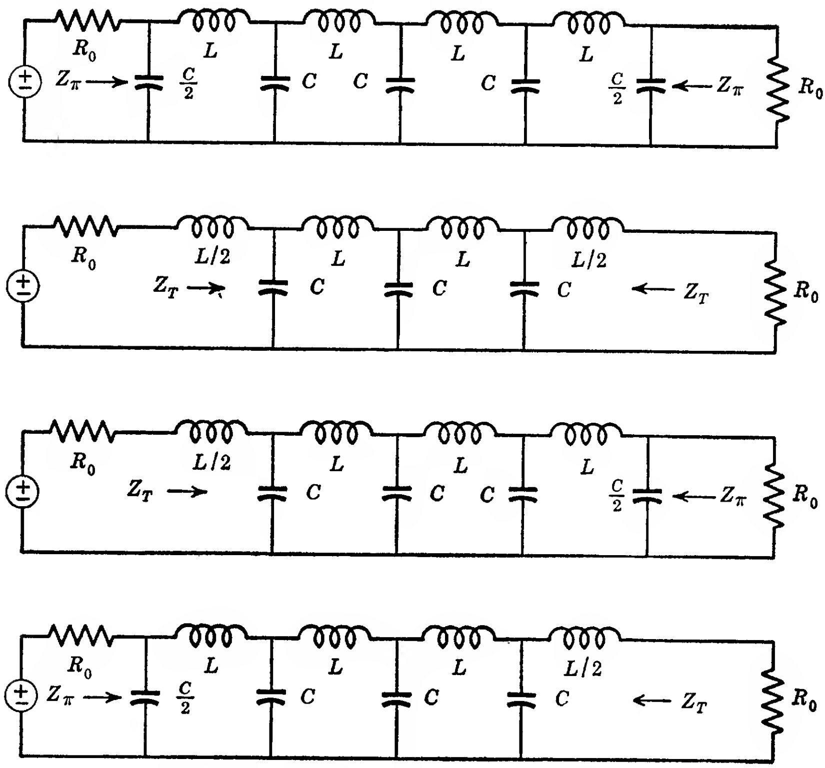


Fig. 8.5. Examples of constant-k filter chains.

The ideal design assumes that the terminations are exactly the image impedances Z_T or Z_π. We have shown resistors as actual terminations in Fig. 8.5. It behooves us to determine just what Z_T and Z_π are in the pass band in order to see just how well we have succeeded in matching. We have in the pass band

$$\begin{aligned} Z_T &= R_0(1 - \omega^2 LC/4)^{1/2} = R_0[1 - (\omega/B)^2]^{1/2} \\ Z_\pi &= R_0/(1 - \omega^2 LC/4)^{1/2} = R_0/[1 - (\omega/B)^2]^{1/2} \end{aligned} \tag{8.15}$$

which is plotted in Fig. 8.6. It can be seen that the match is perfect only at $\omega = 0$ and becomes progressively worse as ω increases towards $2/(LC)^{1/2} = B$. Therefore we should not expect to get the ideal characteristics but only something approximating them. For example, the ratio of output voltage to source voltage for a single T or pi section terminated in R_0 can be examined to see that it gives the maximally flat three-pole transfer function which has attenuation and phase characteristics only similar to the ideal image function.

For a multisection filter, the behavior is more complex. Let us study a multisection symmetric constant-k filter having R_0 terminations with

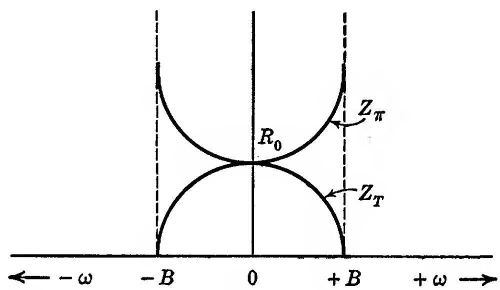


Fig. 8.6. T and pi image impedances in the pass band.

the aid of the insertion ratio involving mismatch factors of eq. 7.60. It will be assumed for the moment that incidental loss in the coils and capacitors is large enough so that the interaction factor σ remains approximately at unity. Then, eq. 7.60 becomes

$$\frac{i_{20}}{i_2} = e^{jn\beta}\left[\frac{[(1 - \omega^2/B^2)^{1/2} + 1]^2}{4(1 - \omega^2/B^2)^{1/2}}\right] \tag{8.16}$$

which is valid for either T or pi sections, where n is the total number of sections, and in which the multiplier of the exponential is the deviation from the ideal. This expression is valid only in the pass band where ideally α is zero. The multiplier of eq. 8.16 is shown in Fig. 8.7. At the edge of the pass band, the correction term becomes infinite, implying infinite attenuation, although the presence of the interaction factor prevents this from being correct. Nevertheless, the correction term does differ markedly from unity as the edge of the pass band is approached, which causes the attenuation to increase somewhat with frequency in the pass band rather than remaining at zero as when terminations are the ideal image impedances.

Let us see what happens to this correction term if we terminate in resistances something different than R_0. If we actually terminate in

hR_0 for pi sections adjacent to the terminations or in R_0/h if we use T sections next to the terminations, then the insertion ratio (neglecting the interaction factor) becomes

$$\frac{i_{20}}{i_2} = e^{jn\beta}\left[\frac{[h(1-\omega^2/B^2)^{1/2}+1]^2}{4h(1-\omega^2/B^2)^{1/2}}\right] \tag{8.17}$$

The correction factor for $h = (2)^{1/2}$ is sketched as a dashed line in Fig. 8.7. Since the dashed curve varies less *on the average* than does the

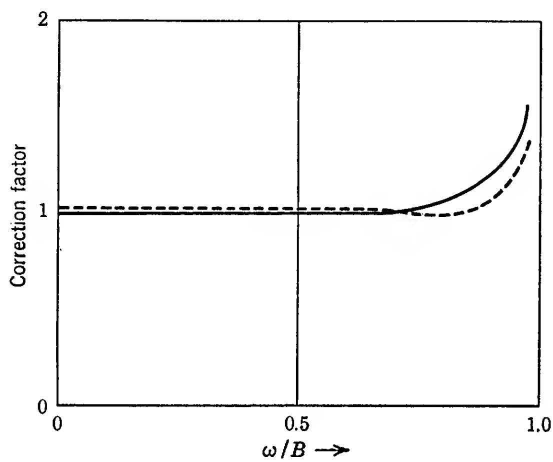


Fig. 8.7. Correction factor for mismatch, neglecting interaction.

solid curve in the pass band (for $h = 1$), we conclude that the use of a resistance somewhat larger than R_0 for pi sections or somewhat less than R_0 for T sections is warranted in many instances. These modified values of terminating resistances are in fact a kind of average image impedance in the pass band as can be seen by inspecting Fig. 8.6.

Now let us study the interaction factor. Again assuming R_0 terminations, we get

$$\sigma = 1 - \left(\frac{(1-\omega^2/B^2)^{1/2}-1}{(1-\omega^2/B^2)^{1/2}+1}\right)^2 e^{-2n(\alpha+j\beta)} \tag{8.18}$$

For α appreciable and for a large number of sections n, σ is little different from unity. However, if we assume the worst case where $\alpha = 0$, the magnitude of σ can be seen to ripple in the pass band a number of times depending upon the number n. By expanding the exponential of eq. 8.18 into sine and cosine terms and finding the magnitude of σ for

$\alpha = 0$, we get

$$|\sigma| = \left[1 + \left(\frac{(1 - \omega^2/B^2)^{1/2} - 1}{(1 - \omega^2/B^2)^{1/2} + 1}\right)^2 (1 - 2\cos 2n\beta)\right]^{1/2} \tag{8.19}$$

The behavior of $|\sigma|$ is roughly indicated in Fig. 8.8 as a function of frequency for n fairly large. (A simple phasor diagram of eq. 8.18 helps in visualizing the behavior of the interaction factor.)

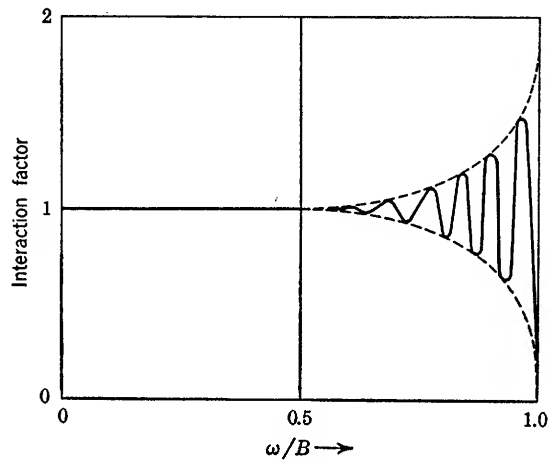


Fig. 8.8. Behavior of the interaction factor.

The combined effect of mismatch and interaction factors is to give an output to input voltage ratio (transfer function) for a multistage filter having R_0 terminations something like that shown in Fig. 8.9*a*. The curve of Fig. 8.9*b* applies when the image match is perfect.

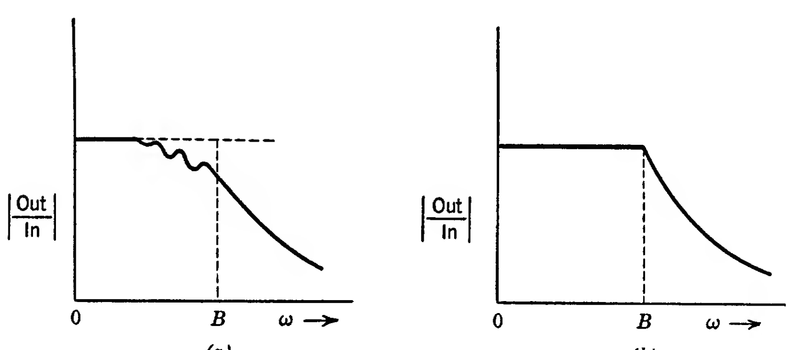


Fig. 8.9. Typical transfer function with mismatch and interaction.

As in the case of the correction factor of eq. 8.17, the effects of the interaction factor are reduced with terminations different from R_0; the ripples are reduced near the band edge and are spread more uniformly over the pass band.

The fact that other than R_0 terminations may be desirable has one rather interesting result. That is, if at one end of a filter a T section is used and at the other a pi section (with an intervening L section some place to maintain an image match) as in the last two filters of Fig. 8.5, a slight amount of impedance transformation can be obtained. If $h = (2)^{1/2}$, the ratio of source and load impedances will be 2. This will result in a perfect image (and conjugate) match at some finite frequency in the pass band. However, there can of course then be no match at $\omega = 0$. The result is a transfer function that is slightly peaked in the pass band.

The ability to terminate in something other than R_0 has yet a further interesting result. In a "correctly" terminated filter (that is, in R_0), the gain characteristic is something of a cross between a Chebyshev and a maximally flat function, depending on the number of sections in the filter. For terminations larger than R_0 for pi sections and smaller than R_0 for T sections (adjacent to the terminations), it is evident that the gain characteristic becomes flatter in the pass band with perhaps a more uniformly rippled nature; in this manner, we can design a constant-k filter with some of the attributes of a Chebyshev filter. On the other hand, for terminations smaller than R_0 for pi sections and larger than R_0 for T sections, the gain characteristic becomes smoother and drops off with frequency more gradually; in this manner, we can get a filter with some of the attributes of a linear-phase filter. For pi sections adjacent to the terminations (which gives a shunt capacitance as a leading element) and for a given bandwidth, the manner in which the product of source resistance and leading capacitance varies with the type of gain characteristic is quite clear; large values of this product go with Chebyshev-type characteristics, and smaller values with linear-phase-type characteristics.

8.3 *m*-derived filters

Two properties of the simple low-pass filter described in the preceding section can in some cases be improved. First, it would be most desirable to invent some L section to be placed between a constant-k filter and a resistive load so that the match between filter and load could be improved. Such a section would be used as indicated in Fig. 8.10, in which the proper constant-k impedance is seen looking into the side of the end section connected to the usual constant-k filter, whereas the

impedance seen looking into the side of the end section connected to the image-impedance load is a much better approximation to a constant resistance in the pass band. In Fig. 8.10, Z_T and Z_π are the same as in the usual constant-k filter, whereas Z_{Tm} and $Z_{\pi m}$ are more nearly constant resistances in the pass band. Then, when a resistor is employed as a termination, the match will be much better than with the simple

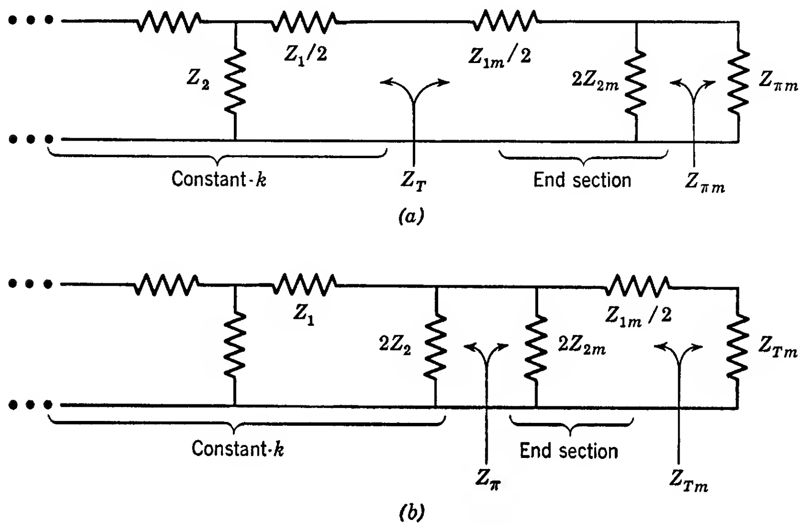


Fig. 8.10. Use of m-derived sections.

constant-k filter and consequently the ideal image characteristics will be more closely realized.

Second, it would be advantageous if we could build a type of T or pi section that furnishes zeros of the transfer function on the $j\omega$ axis so that in the stop band we could get frequencies at which the attenuation becomes infinite. It will be necessary that such sections match the usual Z_T or Z_π so that they can be used in conjunction with the ordinary constant-k filter section. Then, it should be possible to approximate the brick-wall transfer function much better outside the pass band.

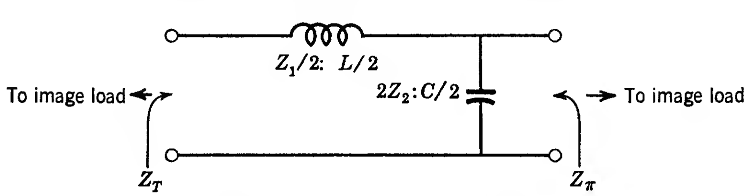


Fig. 8.11. The constant-k L section.

It turns out that m-derived sections give both of the characteristics just named. This new type of network is obtained by manipulating the equations and hoping for good results. We shall go immediately to the proper manipulation by working on an ordinary constant-k L section.

Consider the constant-k L section of Fig. 8.11. The image impedances are

$$\begin{aligned} Z_T &= (z_{10}z_{1s})^{1/2} \\ Z_\pi &= (z_{20}z_{2s})^{1/2} \end{aligned} \tag{8.20}$$

We wish here to manipulate the L section so that Z_T remains invariant but so that Z_π is changed. This can be done in many ways if a new z_{10} and a new z_{1s} are defined as

$$\begin{aligned} z_{1s}' &= F(m)z_{1s} \\ z_{10}' &= z_{10}/F(m) \end{aligned} \tag{8.21}$$

where $F(m)$ is some function of the parameter m. It should be evident that Z_T has not been changed by this manipulation.

Let us take the simplest case; that is, let us assume that $F(m) = m$. Then

$$\begin{aligned} z_{1s}' &= mz_{1s} = mj\omega L/2 = Z_{1m}/2 \\ z_{10}' &= z_{10}/m = j\omega L/2m + 2/j\omega Cm = 2Z_{2m} + Z_{1m}/2 \end{aligned} \tag{8.22}$$

in which $Z_{1m}/2$ and $2Z_{2m}$ are the impedances for the arms of the "derived" L section. Solving this set of equations, we get

$$\begin{aligned} \frac{Z_{1m}}{2} &= \frac{j\omega mL}{2} \\ 2Z_{2m} &= \frac{j\omega L(1 - m^2)}{2m} + \frac{2}{j\omega mC} \end{aligned} \tag{8.23}$$

The original constant-k section and the new derived L section are shown in Fig. 8.12. The impedance Z_T for the original and derived L

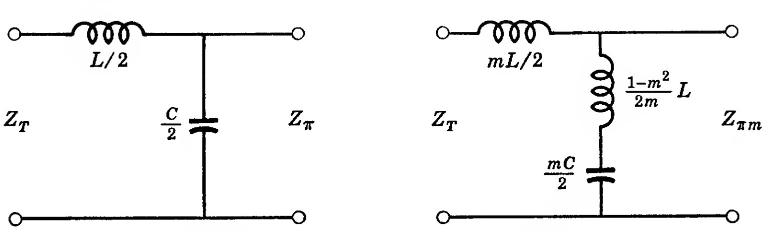


Fig. 8.12. The series-derived end section.

sections remains unchanged because we made it so. However, Z_π has been changed. Its value in the pass band is found to be

$$Z_{\pi m} = Z_\pi[1 - \omega^2(1 - m^2)/B^2] \qquad (8.24)$$

in which Z_π is the usual constant-k filter impedance. It should be remembered that Z_π increases with frequency. However, because the term in brackets in eq. 8.24 decreases with frequency, it is possible to make Z_π fairly constant with frequency (and it is still purely resistive in the pass band) up to frequencies fairly near the edge of the pass band.

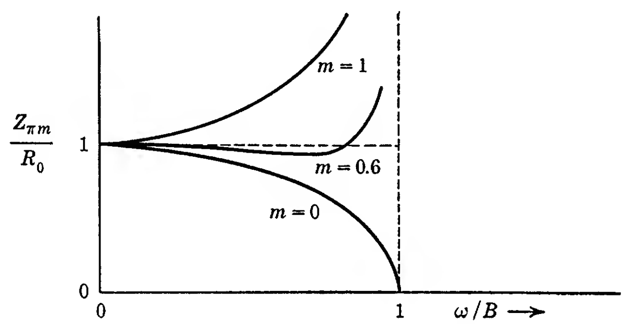


Fig. 8.13. Variation of $Z_{\pi m}$ with m.

Note that for $m = 1$ all the preceding equations and circuits for the derived section reduce to the ordinary constant-k situation, as they should.

$Z_{\pi m}/R_0$ is sketched as a function of ω/B for a few values of m in Fig. 8.13. In particular, $Z_{\pi m}$ is fairly close to being constant in the pass band for $m = 0.6$. The curve for $m = 1$ gives the ordinary constant-k impedance.

It follows that a good way to terminate the constant-k filter is to use an m-derived L section (m about 0.6) next to the source and load resistances R_0 as indicated in Fig. 8.14. Of course, if an extra constant-k L section is used somewhere in the filter, a series-derived section (the one just discussed) can be used to match one termination, and a shunt-derived section to match the other. (We shall take up the shunt-derived section shortly.)

It is now of interest to study the attenuation characteristics of the derived filter. For this, we shall use two m-derived L sections back to back to give a full section as in Fig. 8.15. Now, the image impedance looking into the section from either end is the ordinary constant-k impedance Z_T. Therefore, such a section can replace a constant-k section in a constant-k filter chain without disturbing the image match. Also,

note the presence of the series-resonant circuit in the m-derived section. Signals passing through the filter at the resonant frequency will be

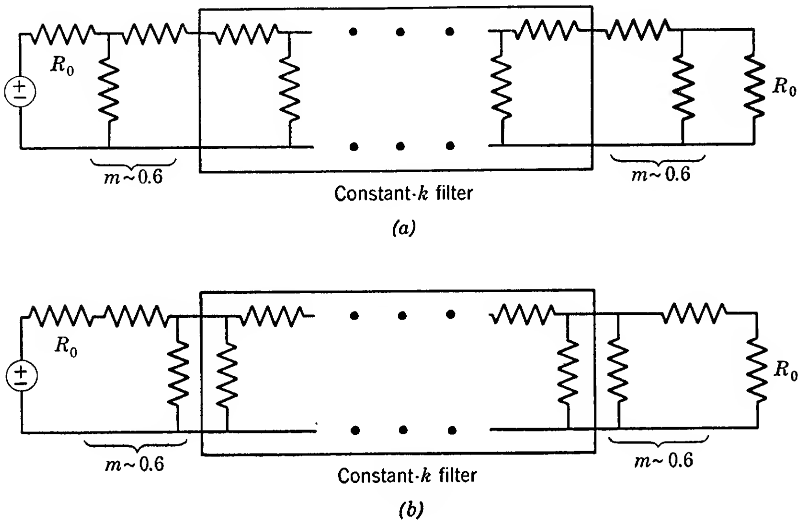


Fig. 8.14. Use of derived sections for matching. (*a*) Series-derived type. (*b*) Shunt-derived type.

short-circuited to ground, resulting in an infinite attenuation at the frequency

$$\omega_\infty = \frac{B}{(1 - m^2)^{1/2}} \tag{8.25}$$

where $B = 2/(LC)^{1/2}$ is the constant-k bandwidth. Since we always

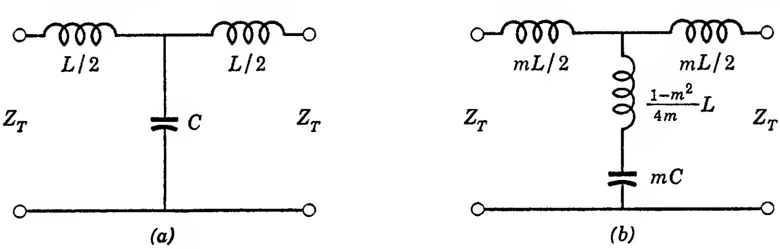


Fig. 8.15. The full series m-derived section.

have $m \leq 1$, the frequency of infinite attenuation always falls in the stop band of the filter.

The pass band still has $\alpha = 0$ and the stop band has an attenuation

given by

$$\cosh \alpha = \pm\left(-1 - \frac{Z_{1m}}{2Z_{2m}}\right) = \pm\left(\frac{2\omega^2 m^2/B^2}{1 - \omega^2(1 - m^2)/B^2} - 1\right) \quad (8.26)$$

where the positive sign applies for $B < \omega < \omega_\infty$ and the negative sign for $\omega > \omega_\infty$. The variation of α with frequency obtained from this equation is sketched in Fig. 8.16 for various values of m. These curves refer to a full section. For the half-section, the values of α are just half those shown in Fig. 8.16. If half-sections are used at both ends of a

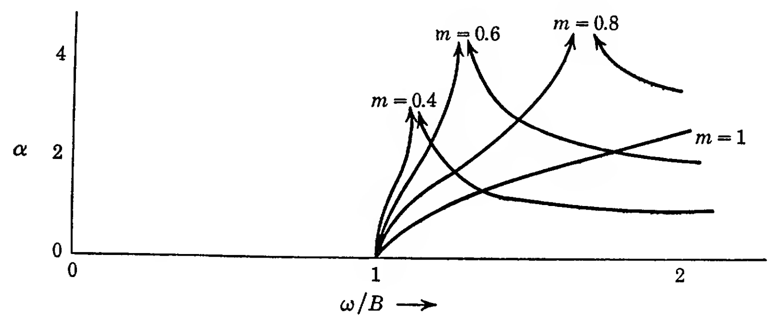


Fig. 8.16. Attenuation of the full m-derived section.

filter, the two half-sections taken together are equivalent to a full section.

Note that, in order to get high attenuation just outside the pass band, a section with a small value of m is indicated. However, sections with small m give low attenuation at high frequencies. In building up a composite filter, m-derived sections are indicated in order to make the attenuation large near the edge of the pass band (as well as to give a better resistive match). In addition, constant-k sections may be necessary to give high attenuation at high frequencies. If we let $\omega \to \infty$ in eq. 8.26, we get the high-frequency attenuation of an m-derived full section as

$$\lim_{\omega \to \infty} \cosh \alpha = 1 + 2m^2/(1 - m^2) \quad (8.27)$$

Up to now, we have studied the series-derived section in which Z_T is maintained unchanged and Z_π is modified. The reverse of this is the shunt-derived section, the full section of which is a pi network. We shall leave the derivation of the pertinent equations to the reader and merely set down the results here. Half and full shunt-derived sections are shown in Fig. 8.17. The same attenuation curves apply for either shunt- or series-derived sections. Also, the same equation for ω_∞ applies.

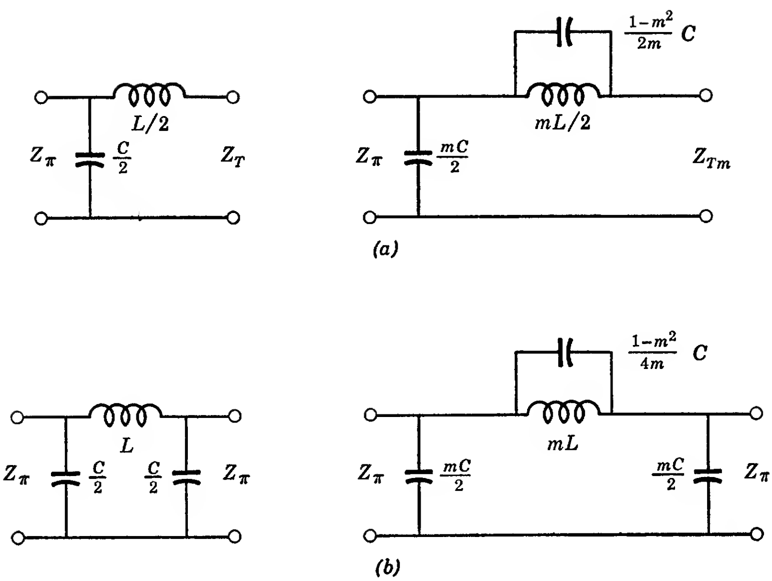


Fig. 8.17. Shunt-derived sections.

The only difference is Z_{Tm} which, for the shunt-derived section, is given by

$$Z_{Tm} = \frac{Z_T}{1 - \omega^2(1 - m^2)/B^2} \tag{8.28}$$

which plots as shown in Fig. 8.18. As before, the impedance Z_{Tm} is

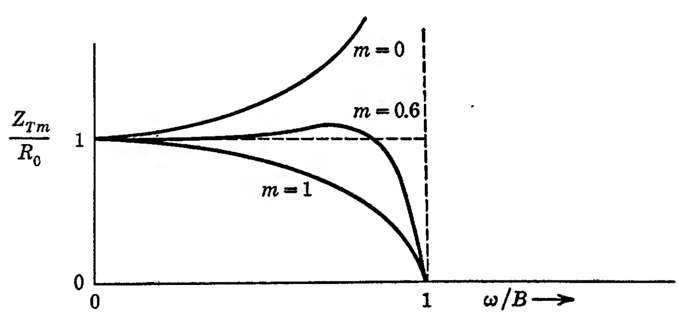


Fig. 8.18. Variation of Z_{Tm} with m.

purely resistive in the pass band and is fairly constant at R_0 for m about 0.6. The value $m = 1$ applies to the simple constant-k filter.

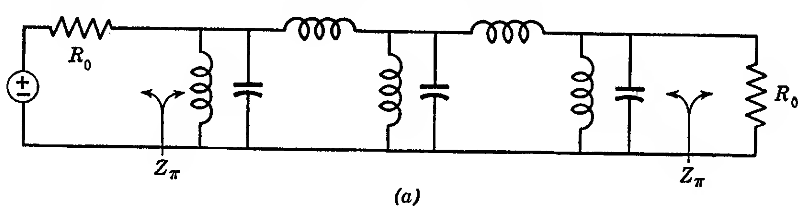
We originally took the function $F(m)$ of eqs. 8.21 as simply m. Of course, we could have tried more complicated functions of m to derive

more complex m-derived networks, some of which give a better resistive match than the simple m-derived type. These more complex networks are referred to as "multiply-derived" (m-m) types. Although they are capable of giving excellent matching characteristics, their complexity is often too great to be practical except in very special cases.

The m-derived section was obtained here in reference to the simple low-pass filter. The derivation can be made more general than we have done so that the various equations and curves are expressed in terms of the general ladder impedances Z_1 and Z_2 and the ratio $Z_1/4Z_2$ instead of the frequency ω. However, the most practical circuits have been treated specifically (including band-pass and other equivalents of these circuits) and the more general relations are seldom needed.

8.4 An image-matched band-pass filter

In this section, we shall derive the design equations for a type of band-pass filter that does not have a simple low-pass equivalent. The purpose of this section is not only to describe an important filter but



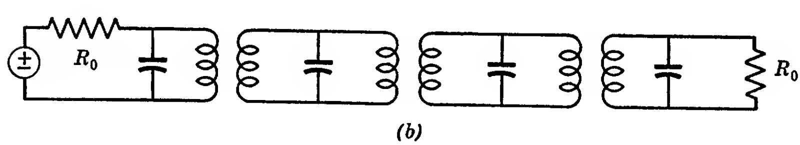


Fig. 8.19. A band-pass image-matched filter.

also to indicate how the image relations apply to other filters than the simple low-pass one.

Intuitively, it might be expected that the ladder filter of Fig. 8.19a gives a band-pass behavior. This network can be realized in the form of a series of transformers as in Fig. 8.19b. A very nice feature of this band-pass filter is that it requires fewer circuit elements per section than does the band-pass equivalent of the simple constant-k low-pass filter.

Figure 8.20 shows the L and full pi sections of the filter to be studied. With $Z_1 = j\omega L_1$ and $Y_2 = j\omega C_2 + 1/j\omega L_2$, we get

$$Z_1/4Z_2 = (L_1/4L_2)(1 - \omega^2 L_2 C_2) \tag{8.29}$$

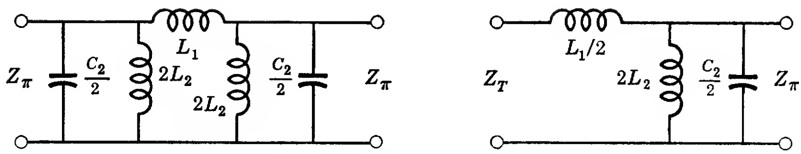


Fig. 8.20. Pi and L sections of the band-pass filter.

which is sketched as a function of frequency in Fig. 8.21. Here are shown the pass band, lower and upper band edges ω_1 and ω_2, bandwidth

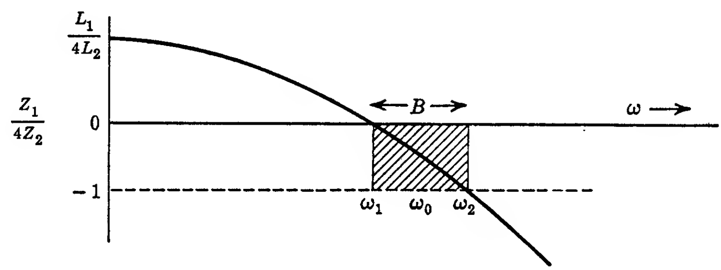


Fig. 8.21. $Z_1/4Z_2$ for the band-pass filter.

B, and center frequency ω_0. By means of eq. 8.29, we easily find

$$\begin{aligned} \omega_1 &= 1/(L_2C_2)^{1/2} \\ \omega_2 &= \omega_1(1 + 4L_2/L_1)^{1/2} \\ B &= \omega_2 - \omega_1 = \omega_1[(1 + 4L_2/L_1)^{1/2} - 1] \end{aligned} \tag{8.30}$$

The pi image impedance is found to be

$$Z_\pi = (z_{20}z_{2s})^{1/2} = \left[\frac{4L_2^2\omega^2}{[(\omega/\omega_1)^2 - 1][(\omega_2/\omega_1)^2 - (\omega/\omega_1)^2]}\right]^{1/2} \tag{8.31}$$

which is purely real in the pass band and is a curved function of frequency as is indicated by Fig. 8.22. It is most logical to define the

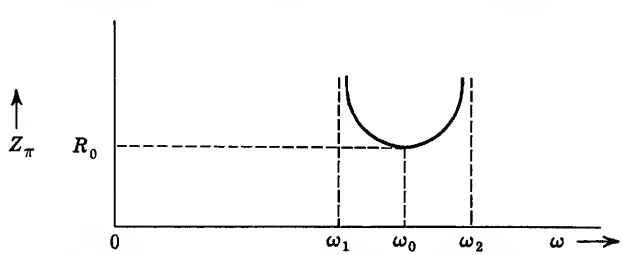


Fig. 8.22. Image impedance in the pass band.

center frequency at the minimum of this curve and to define also at this frequency the characteristic resistance R_0. Setting the derivative of eq. 8.31 equal to zero and thereby determining ω_0, we get

$$\omega_0 = (\omega_1\omega_2)^{1/2} = \omega_1(1 + 4L_2/L_1)^{1/4} \tag{8.32}$$

With this value of frequency in eq. 8.31, Z_π becomes R_0 by definition. The result is

$$R_0 = 2\omega_1{}^2L_2/B \tag{8.33}$$

We now have enough equations to determine all the circuit elements given only R_0, ω_0, and B. We find

$$C_2 = 2/BR_0 \tag{8.34}$$

from which the rest of the element values follow.

The character of the image impedance Z_π in the pass band is similar to that for constant-k pi networks. Therefore it can be expected that terminating the filter in a resistance somewhat larger than R_0 improves the behavior (in one sense).

8.5 Constant-k matching networks

It will be recalled that Z_T and Z_π for the constant-k L section vary inversely with respect to one another as

$$\begin{aligned} Z_T &= R_0[1 - (\omega/B)^2]^{1/2} \\ Z_\pi &= R_0/[1 - (\omega/B)^2]^{1/2} \end{aligned} \tag{8.35}$$

where $B = 2/(LC)^{1/2}$ is the bandwidth and where both Z_T and Z_π are purely resistive for $\omega < B$.

Suppose we connect source and load resistances R_1 and R_2 together with a single L section as in Fig. 8.23. At some finite frequency ω less

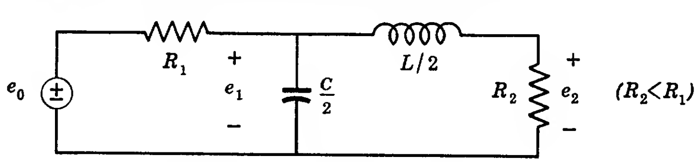


Fig. 8.23. A simple matching network.

than B, let us adjust matters so that $R_1 = Z_\pi$ and $R_2 = Z_T$. These relations placed in eqs. 8.35 result in

$$\begin{aligned} R_2 &= R_0[1 - (\omega_0/B)^2]^{1/2} \\ R_1 &= R_0/[1 - (\omega_0/B)^2]^{1/2} \\ R_2/R_1 &= 1 - (\omega_0/B)^2 \end{aligned} \tag{8.36}$$

Therefore, a ω_0 can be selected so that we may match a source resistance R_1 at the frequency ω_0 to a lower load resistance R_2. The ratio e_2/e_0 will be $(R_2/4R_1)^{1/2}$ (conjugate match) at the frequency ω_0. At $\omega = 0$, e_2/e_0 will be less than $(R_2/4R_1)^{1/2}$. In fact, we might expect the magnitude of e_2/e_0 to vary with frequency as shown in Fig. 8.24. If ω_0 is very small, then resistors that differ but little can be matched. However, the bandwidth over which the match is good is quite large; it is B

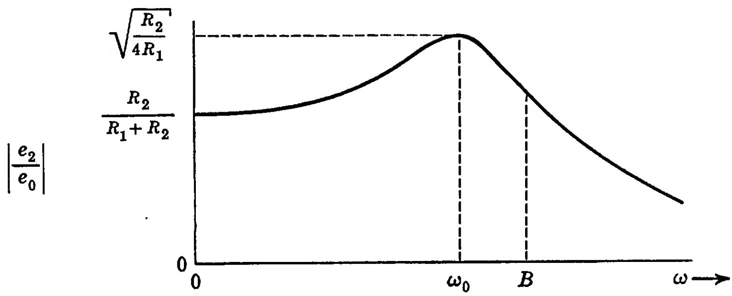


Fig. 8.24. Transfer function of the matching network.

if $\omega_0 = 0$ because the filter is then the simple low-pass L section. As ω_0 increases, which implies that the ratio of R_2 to R_1 decreases, the match is still good at ω_0 although the bandwidth over which the match is good decreases because of the more rapid variations of Z_T and Z_π with frequency as $\omega = B$ is approached. Therefore, if a wide-band match is desired between quite different resistances, it is best to use two or more L sections with each section transforming the impedance by a ratio of less than R_2/R_1. It appears most efficient to transform in equal

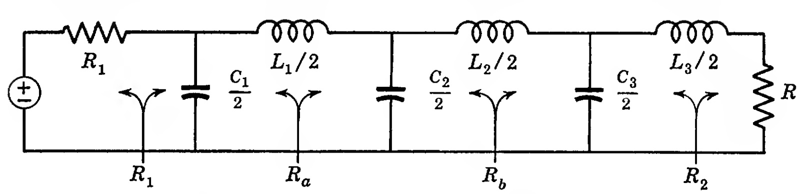


Fig. 8.25. A multisection matching network.

percentage steps. A several-section matching network is shown in Fig. 8.25, in which

$$R_2/R_b = R_b/R_a = R_a/R_1 \tag{8.37}$$

When carried to an extreme where each L section transforms impedances differing by a very small amount, the multisection filter becomes the circuit equivalent of the tapered transmission line.

The theorem of reciprocity permits the source to be placed in the branch containing R_2 with the load becoming R_1 so that an L section can match a low source resistor to a higher load resistor.

It is also possible to employ high-pass L sections consisting of series capacitors and shunt inductors as matching networks. Then, the frequency of the match is higher than the low-frequency cutoff. The details of high-pass matching networks are left to the problems. (In any event, the high-pass matching filter can be derived from the low-pass filter using the standard low-pass transformation.)

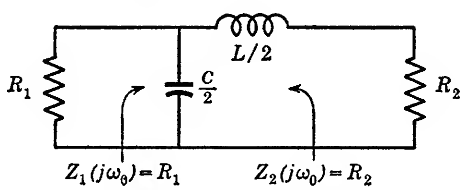


Fig. 8.26. Circuit for deriving parameters of a matching network.

L-section matching networks can be designed in a much more prosaic manner by setting the general expressions for the input and output impedances equal to the desired resistances at the frequency ω_0 as indicated in Fig. 8.26. The parameters of symmetric pi-section matching networks having series L and shunt C can be obtained in a similar manner, as can other networks.

8.6 Constant-resistance lattice networks

All the symmetric filters about which we have spoken up to now can be derived from lattice filters. Because the symmetric lattice is so general, there exist numerous lattice filters (only some of which can be decomposed to ladder form) that are best designed as lattices.

The image impedance of a symmetric lattice network is

$$z_1 = z_2 = (Z_a Z_b)^{1/2} \tag{8.38}$$

Let us first study lattices for which the image impedance is a constant resistance R_0 for *all* frequencies. Then

$$Z_a Z_b = R_0^2 \tag{8.39}$$

from which we see that the lattice impedances Z_a and Z_b must be reciprocal impedances with respect to the resistance R_0. The most impressive thing about constant-resistance networks is that, when ter-

minated in a resistor R_0, the actual behavior of the filter is *exactly* as specified by the image parameters.

From Chap. 7

$$(e^{\phi})_{\text{Lat}} = \frac{1 \pm (Z_a/Z_b)^{1/2}}{1 \mp (Z_a/Z_b)^{1/2}} \tag{8.40}$$

where the sign is dependent upon which impedance is defined as Z_a and which as Z_b.

Using eq. 8.39 in eq. 8.40 for constant-resistance networks

$$(e^{\phi})_{\text{Lat}} = \frac{R_0 \pm Z_a}{R_0 \mp Z_a} \tag{8.41}$$

Let us study the constant-resistance lattice having purely reactive arms. Then, $Z_a = jX_a$ and $Z_b = jX_b = R_0{}^2/jX_a$ and the poles of Z_a are the zeros of Z_b, and so forth. Then

$$(e^{\phi})_{\text{Lat}} = \frac{R_0 + jX_a}{R_0 - jX_a} = 1 \underline{/2 \tan^{-1} \frac{X_a}{R_0}} \tag{8.42}$$

which has unity magnitude for all frequencies. Thus, *the constant-resistance lattice with reactive arms is an all-pass network* which provides a phase shift but no change in magnitude to signals passing through it. Such networks are useful for correcting the phase shift of some other filter without affecting the magnitude characteristics. Impedance Z_a has a certain collection of p-z along the $j\omega$ axis. At a zero of Z_a, the angle $\tan^{-1} X_a/R_0$ is an even multiple of $\pi/2$ radians, and the total phase shift given by the lattice is twice this. At a pole of X_a, $\tan^{-1} X_a/R_0$ is an odd multiple of $\pi/2$ radians. To design an all-pass lattice network, p-z are placed as desired as long as they obey the separation property, and so forth. The lattice arms are then realized, perhaps as one of the four canonical reactance networks.

Unfortunately, the all-pass lattice network cannot be decomposed into an unbalanced ladder but must remain essentially a bridge network. Also, it is not possible to build an ideal constant-resistance network, whether or not the lattice arms are reactive, if there exist nonnegligible shunt capacitances across the source and load impedances.

We have seen that a constant-resistance lattice with reactive arms is an all-pass network. Let us now see if we can find one that is not an all-pass network. Clearly, *we must be willing to accept resistance in the lattice arms in order to do this.* Still, Z_a and Z_b must be reciprocal impedances. A simple constant-resistance lattice with resistance in the

arms is shown in Fig. 8.27*a*. The reader can check to see that the arms are reciprocal and that the image impedance is R_0. This lattice can be decomposed into the bridged-T unbalanced network of Fig. 8.27*b*, which

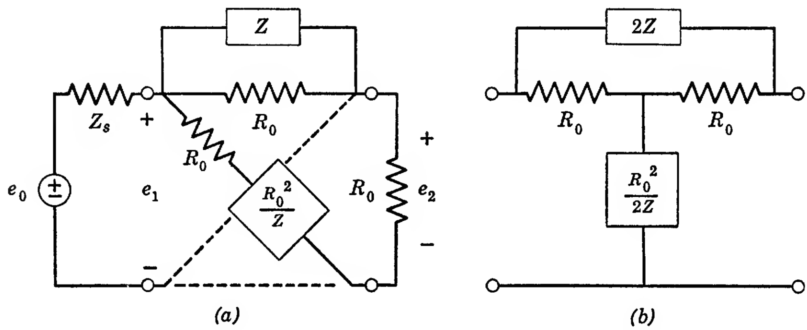


Fig. 8.27. A constant-resistance lattice and equivalent.

is a very practical circuit. The attenuation function of this constant-resistance lattice is

$$\frac{e_1}{e_2} = \frac{R_0 + Z_a}{R_0 - Z_a} = \frac{R_0 + 2Z}{R_0} \tag{8.43}$$

Because the network is constant resistance, we need not employ an R_0 source resistance, although we must use the proper R_0 load resistance. In fact, we can use any source impedance Z_s and still calculate the ratio

$$e_1/e_0 = R_0/(R_0 + Z_s) \tag{8.44}$$

Combining eqs. 8.43 and 8.44, we get the transfer function as

$$\frac{e_2}{e_0} = \left(\frac{R_0}{R_0 + Z_s}\right)\left(\frac{R_0}{R_0 + 2Z}\right) \tag{8.45}$$

As a practical example, let $Z_s = R_s$ and

$$Z = pL + 1/pC \tag{8.46}$$

which is a simple series-resonant circuit. Let us use n of the resulting bridged-T networks in cascade. The over-all transfer function is then

$$\frac{e_n}{e_0} = \frac{R_0}{R_0 + R_s}\left(\frac{pR_0/2L}{p^2 + pR_0/2L + 1/LC}\right)^n \tag{8.47}$$

which is a band-pass function having n zeros at the origin and n superimposed pairs of complex-conjugate poles. The filter appears in Fig. 8.28.

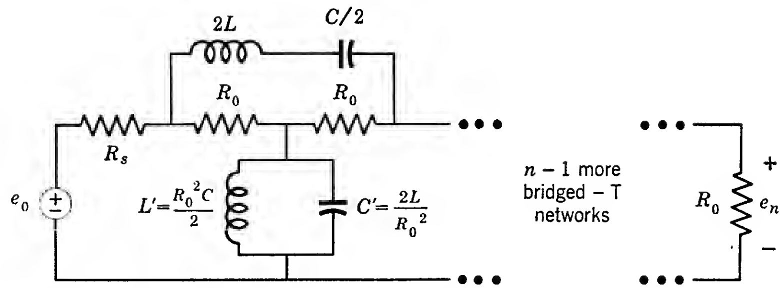


Fig. 8.28. A filter chain of identical bridged-T networks.

8.7 Lossless image-matched lattice filters

Since the lattice is the most general of all symmetric networks, and since there is a lattice equivalent for all the symmetric networks we have studied in this chapter, it is worth while to study the design of networks directly in the form of lattices. It turns out that it is actually simpler in many instances to design a lattice and then decompose it to a ladder (if possible) than to design the filter directly in ladder form.

We shall restrict ourselves to lossless lattice impedances. We are not interested here in the all-pass filter and hence must be willing to settle for something other than a constant-resistance network. The equations of interest are

$$\begin{aligned} z_1 &= z_2 = (Z_a Z_b)^{1/2} \\ \cosh\phi &= \cosh\alpha\cos\beta + j\sinh\alpha\sin\beta = \frac{1 + Z_a/Z_b}{1 - Z_a/Z_b} \end{aligned} \tag{8.48}$$

Since Z_a and Z_b are both reactances, the product Z_aZ_b and the ratio Z_a/Z_a are both purely real. The image impedance is purely real only if Z_a and Z_b are reactances of the opposite types. Also, since eqs. 8.48 are purely real

$$\sinh\alpha\sin\beta = 0 \tag{8.49}$$

which can be satisfied in two ways

$$\begin{aligned} \alpha &= 0: && \text{Pass band} \\ \beta &= 0, \quad \pm\pi: && \text{Stop band} \end{aligned} \tag{8.50}$$

In the pass band, $\cos\beta$ is limited to ± 1. Therefore

$$-1 \le \frac{1 + Z_a/Z_b}{1 - Z_a/Z_b} \le 1 \tag{8.51}$$

which reduces to

$$Z_a/Z_b \leq 0 \tag{8.52}$$

Therefore, whenever Z_a and Z_b are reactances of the opposite types, there will be a pass band and the image impedance will be purely resistive. Whenever Z_a and Z_b are reactances of the same type, there will be a stop band and the image impedance will be purely reactive. These

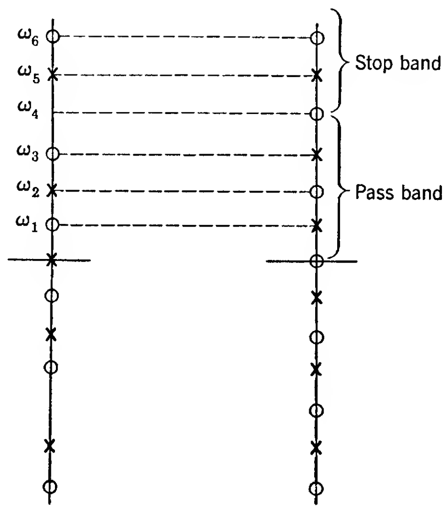


Fig. 8.29. Lattice arm impedances of a low-pass filter.

conditions are somewhat simpler than those relating to T and pi networks.

The condition for infinite attenuation is clearly $Z_a = Z_b$, which is the null condition for a bridge. Consequently, points of infinite loss correspond to a bridge balance.

The p-z of the reactance functions representing Z_a and Z_b are quite evident on the p plane. The poles of Z_a must be the zeros of Z_b and conversely in a pass band, otherwise the reactances cannot be of the opposite types, whereas the poles of Z_a must also be the poles of Z_b in a stop band, or else the reactances cannot be of the same type. A plot of the p-z for Z_a and Z_b for a type of low-pass filter appears in Fig. 8.29. (It does not matter which plot of Fig. 8.29 represents Z_a and which represents Z_b.) Note for this example that impedances Z_a and Z_b change from being opposite types of reactances to being the same type at the frequency ω_4 where a zero of one of the impedances is not matched by a pole of the other impedance. Further, note that by using the reciprocal reactance for one (not both) of the two reactances the char-

acteristics of the filter are inverted so that pass bands become stop bands and conversely.

The square root of the product of Z_a and Z_b for the example of Fig. 8.29 is easily seen to be

$$z_1 = z_2 = (Z_a Z_b)^{1/2} = H \frac{(p^2 + \omega_4{}^2)^{1/2}(p^2 + \omega_6{}^2)}{p^2 + \omega_5{}^2} \tag{8.53}$$

which, for $\omega < \omega_4$, is the image impedance in the pass band, and where H is a real constant multiplier. In this product, it can be observed that all the p-z of the two impedances occurring in the pass band cancel.

It is desirable to make the image impedance, as exemplified by eq. 8.53, approximate a resistance that is constant in the pass band by suitably adjusting the frequencies ω_4, ω_5, and ω_6. This may be done in several ways analogous to that done in approximating the brick-wall transfer function. If eq. 8.53 is made to ripple about an average value, the match using resistive terminations will ripple about the ideal image match and result in a rippled transfer function. Alternately, eq. 8.53 can be made to approximate a constant in the Taylor sense, so that the transfer function with resistive terminations will look something like

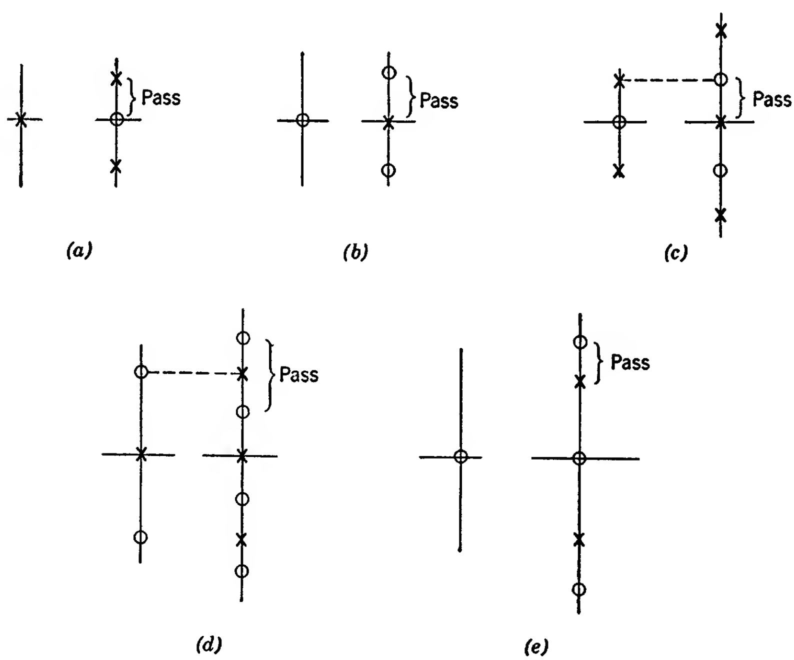


Fig. 8.30. Lattice impedances of several ladder filters.

the maximally flat transfer function. We do not have time to study these various approximating procedures here but must leave them for individual study from the references. (The reader should note that the most efficient use of p-z in maintaining the image impedance constant in the low-pass band, using a given total number of p-z, requires that the lowest frequency pole or zero, other than that at $\omega = 0$, define the edge of the low-pass band.)

A few p-z plots for Z_a and Z_b corresponding to networks treated earlier in this chapter are shown in Fig. 8.30. (Bartlett's bisection theorem has been used to find the representative lattice impedances.) Figure 8.30*a* shows Z_a and Z_b for the lattice equivalent to the simple low-pass constant-k filter. In *b*, the p-z are for the series m-derived section, whereas in *c* the p-z are for the shunt-derived section. Figure 8.30*d* is for the band-pass equivalent of the low-pass constant-k filter and *e* is for the special band-pass filter treated in Sec. 8.4. As will be recalled, reversing p-z of any one of the two impedances of Fig. 8.30 results in the inversion of stop and pass bands. With this minor change, Fig. 8.30 also represents several high-pass and band-elimination filters.

Problems

1. A ladder filter has $Z_1 = pL_1 + 1/pC_1$ and $Z_2 = 1/pC_2$. Plot the variation of $Z_1/4Z_2$ with frequency. Locate the pass and stop bands in terms of element values. Evaluate Z_T and Z_π and plot in the pass band. Define R_0 at some unique point in the pass band. Evaluate α as a function of frequency for both stop and pass bands.

2. A constant-k low-pass filter has an image-impedance load and is driven from a voltage source with a source impedance of zero. Plot the image-attenuation function and compare with that given when the source impedance is also the image impedance.

3. How many constant-k full sections are needed to give an attenuation of 40 decibels at a frequency of 1.2 times the bandwidth? (α is expressed in terms of nepers. The number of decibels is 8.686 times the number of nepers.)

4. Design a three-section constant-k filter using pi sections for a bandwidth of 100 mcs and a full shunt C of 12 $\mu\mu$f. Plot the ideal attenuation function. Plot the actual attenuation function in the pass band assuming the interaction factor σ to be unity for (*a*) R_0 terminations and (*b*) $\sqrt{2}\,R_0$ terminations.

5. Convert the filter of Prob. 4 to a band-pass filter at a center frequency of 50 mcs (with the same bandwidth as in Prob. 4). Use $\sqrt{2}\,R_0$ terminations. What are the lower and upper band edges? Use determinant manipulation at the output node of the network to double the load impedance without changing the input impedance. Sketch the transfer function using the results of Prob. 4 as a guide.

6. Design a two-section (pi) high-pass constant-k filter using R_0 terminations and having $R_0 = 1600\ \Omega$ and a low-frequency cutoff of 10 kcs.

7. Plot the transfer function in the pass band of a low-pass filter (neglect the interaction factor) composed of one pi section and one L section having $B = R_0 = 1$. Use $\sqrt{2}\ \Omega$ at one end of the filter and $1/\sqrt{2}$ at the other to get a 2:1 ratio of resistances.

8. Derive the relations for the shunt m-derived low-pass filter section.

9. Design a low-pass image-matched filter having the attenuation function of Fig. P.9. Use $m = 0.6$ series-derived end sections and use $R_0 = 300\ \Omega$. Use one internal constant-k section and one internal m-derived section.

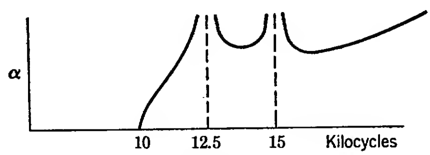


Fig. P.9.

10. Design a ladder filter having pi sections with the following specifications: $R_0 = 600\ \Omega$ terminations, 6 kcs low-pass bandwidth, good image match (use m-derived end sections with m approximately 0.6). The attenuation should rise to infinity at 7 kcs and remain above 40 decibels for all higher frequencies.

11. A constant-k band-pass filter passing frequencies between 20 and 40 mcs is desired with infinite attenuation at frequencies somewhat above and below the band edges and with a good image match. $R_0 = 800\ \Omega$. Design this filter using a single shunt-derived section converted to its band-pass equivalent.

12. A single-section constant-k filter with equal terminations has the form of Fig. P.12. Show that the transfer functions of these networks can always be partially factored so that the locations of the p-z can be determined.

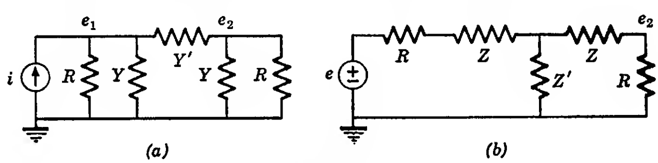


Fig. P.12.

13. Plot the locations of the p-z of a constant-k pi network with R_0 and $\sqrt{2}\,R_0$ terminations.

14. Plot the locations of the p-z of an m-derived pi section with R_0 terminations.

15. The general L network is shown in Fig. P.15a. The series-derived network is indicated in Fig. P.15b. Determine the impedances of the m-derived L section in terms of Z_1 and Z_2.

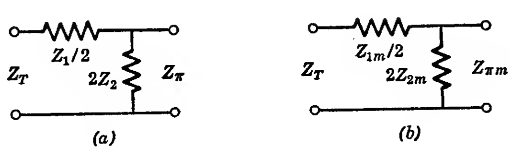


Fig. P.15.

16. Plot Z_T, Z_π, and $Z_{\pi m}$ as functions of $Z_1/4Z_2$ for the networks of Prob. 15.

17. Determine the element values for the network of Fig. 8.19a to have a bandwidth of unity at a center frequency of two. Plot the ideal transfer function for

$R_0 = 1\text{K}$ image-impedance terminations. Plot the actual transfer function (neglecting the interaction factor) for R_0 and $\sqrt{2}\,R_0$ resistive terminations.

18. Design a band-pass network with the circuit of Fig. 8.19*a* for a bandwidth of 4.8 mcs centered at 21 mcs and with $\sqrt{2}\,R_0$ terminations.

19. Convert the network of Prob. 18 to that shown in Fig. 8.19*b*.

20. An L-section matching network is to match 300 Ω to 50 Ω (load) at a frequency of 1.5 mcs. Determine the network, plot the p-z of the transfer function, plot the magnitude of the transfer function, and determine the half-power bandwidth of the match.

21. Two cascaded constant-*k* low-pass half-sections are used to match a 400 Ω source to a 50 Ω load at a frequency of 28.5 mcs. Design the network.

22. Use a two-section matching network in Prob. 20 instead of a one-section network. Plot the magnitude of the transfer function and determine the bandwidth of the match.

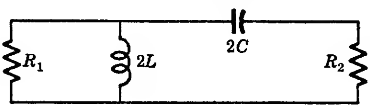


Fig. P.23.

23. Determine the constant-*k* matching equations for the high-pass half-section of Fig. P.23.

24. Determine the equations for designing the matching network of Fig. P.24. Do not use image-matching concepts. Be careful not to cancel identical factors from both sides of an equation.

25. Design an all-pass lattice network to give phase shifts of 0, π, 2π, 3π, and 4π at frequencies of 300, 1000, 1500, 3000, and 6000 cps respectively.

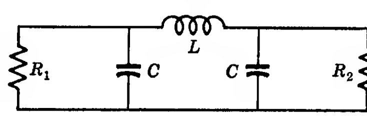


Fig. P.24.

26. Determine Z and develop design criteria for the bridged-T constant-resistance rejection filter.

27. Develop a theory based on p-z plots analogous to those of Fig. 8.29 but applicable to the impedances Z_1 and Z_2 of T and pi sections.

28. A lattice has $Z_a = pL_1 + 1/pC_1$ and $Z_b = 1/pC_2$. Locate the stop and pass bands in terms of element values and sketch the attenuation in the stop band. Attempt to reduce the lattice to a ladder.

29. Repeat Prob. 28 for $Z_a = pL_1 + 1/pC_2$ and $Z_b = pL_2$.

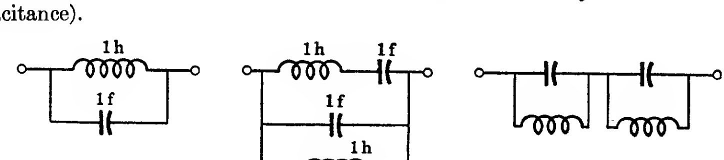
30. Repeat Prob. 28 for $Z_a = pL_1 + 1/pC_2$ and $Z_b = pL_2 + 1/pC_2$.

31. Repeat Prob. 28 for $Y_a = pC_1 + 1/pL_1$ and $Y_b = pC_2$.

32. Repeat Prob. 28 for $Y_a = pC_1 + 1/pL_1$ and $Y_b = 1/pL_2$.

33. Repeat Prob. 28 for $Y_a = pC_1 + 1/pL_1$ and $Y_b = pC_2 + 1/pL_2$.

34. A low-pass *m*-derived filter has the series impedance given by Fig. P.34*a*. When converted to a band-pass filter, the series arm becomes that of Fig. P.34*b*. Determine the more practical equivalent of Fig. P.34*c* in which each inductor is shunted with a capacitor (part of which can be furnished by unavoidable stray capacitance).



(a) (b) (c)

Fig. P.34.

9

The Circuit Representation of Vacuum Tubes

The characteristics of a vacuum tube are described with a set of empirically derived curves. The exact description of these curves can be obtained with a Taylor series. Since more than two variables are involved, this series contains at least two first partial derivatives, several second partials, a quite large number of third partials, and so on. The assumption of linearity disregards all partial derivations above the first. To carry through calculations even as far as the second partials is almost hopeless except in very simple cases. Thus we are limited in calculations to the linear terms of the Taylor series. If we wish to go further, graphical methods are all that are generally practical.

However, the linear assumption is entirely justified whenever the magnitudes of sine-wave voltages and currents are small. Unless a vacuum-tube circuit behaves in a highly nonlinear manner, the assumption of linearity is at least a good first approximation.

The main advantage of the linear approximation is that a vacuum tube can be described in terms of a voltage or current generator and a few other circuit elements. Then the tube can be considered part of some larger network and straightforward network analysis and design carried out with the aid of modern network theory.

In this chapter, we shall study some of the basic properties of vacuum tubes and some of the phenomena closely associated with their use. Later we shall study them in practical systems.

9.1 Simple triode equivalent circuits

A triode is a three-element electron tube containing first a cathode which acts as a source of electrons. Usually the cathode is indirectly heated with a filament which does not play a direct part in the mechanism of operation within the tube. The triode also has a grid which controls the flow of electrons and a plate which acts as a collector of electrons. We shall not consider the phenomena going on inside the triode, but only its properties as a three-terminal network element.

The electrons in a triode flow from cathode to plate and virtually none are intercepted by the grid as long as the grid is negative with respect to the cathode (as we shall always assume to be the case). Since we are studying the triode as a circuit element, the conventional direction of current flow is implied, that is, current flows from plate to cathode.

The basic characteristics of a triode can be displayed with a set of characteristic curves. Although these may be drawn in a variety of ways (such as changing the axes), all the required information is con-

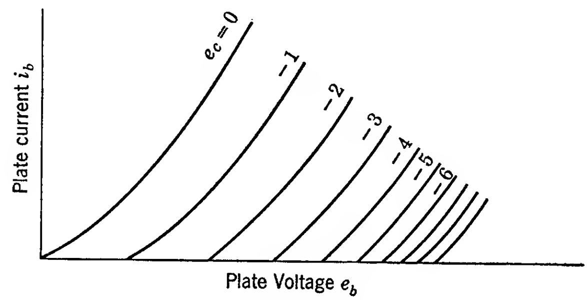


Fig. 9.1. Triode plate characteristics.

tained in the plate-voltage (e_b) plate-current (i_b) curves, a representation of which is given in Fig. 9.1. *All voltages are with respect to the cathode voltage e_k.*

The assumption of linearity replaces these curves with a set where the grid-voltage curves are all straight lines and have a constant separation. Since the separation of the actual curves varies with the grid-to-cathode voltage e_c, a linear approximation made accurate at low negative values of e_c will not agree with one made for high negative values. Two typical approximations are indicated in Fig. 9.2.

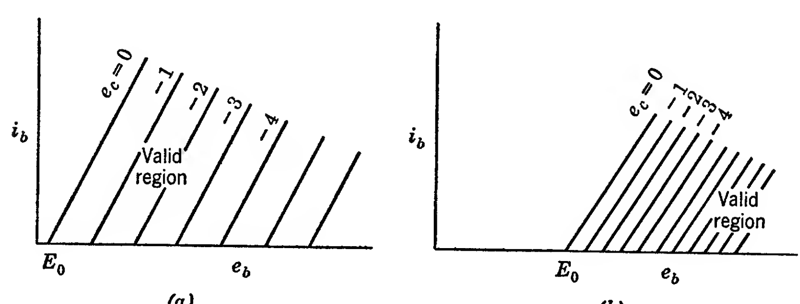


Fig. 9.2. Linear approximation of triode characteristics.

It should be clear that the plate current i_b depends upon two variables, the plate voltage e_b and the grid voltage e_c, both with respect to the cathode. Consequently, the Taylor series linear approximation is

$$i_b = C + \frac{\partial i_b}{\partial e_c} e_c + \frac{\partial i_b}{\partial e_b} e_b \tag{9.1}$$

From the linearized curves at $e_c = i_b = 0$, the plate voltage is E_0. Equation 9.1 then yields

$$C = -\frac{E_0}{(\partial e_b)/(\partial i_b)} \tag{9.2}$$

Both partial derivatives have the dimensions of conductance, which leads us to define two parameters as

$$g_m = \text{Transconductance} = \frac{\partial i_b}{\partial e_c} = \text{Lim}\left(\frac{\Delta i_b}{\Delta e_c}\right)_{e_b \text{ constant}} \tag{9.3}$$

$$r_p = \text{Plate resistance} = \frac{\partial e_b}{\partial i_b} = \text{Lim}\left(\frac{\Delta e_b}{\Delta i_b}\right)_{e_c \text{ constant}} \tag{9.4}$$

where the Δ quantities represent small increments which allow r_p and g_m to be obtained graphically from the characteristic curves.

We can now express eq. 9.1 as

$$i_b = -\frac{E_0}{r_p} + g_m e_c + \frac{e_b}{r_p} \tag{9.5}$$

If this equation is multiplied by r_p, a dimensionless parameter μ is obtained

$$\mu = \text{Amplification factor} = g_m r_p \tag{9.6}$$

and eq. 9.5 takes the form

$$\mu e_c - E_0 - i_b r_p = e_b \tag{9.7a}$$

$$\mu(e_v' - e_k') - E_0 - i_b r_p = e_a' - e_k' \tag{9.7b}$$

where eq. 9.7*b* introduces the total cathode-to-ground voltage e_k', the total grid-to-ground voltage e_v', and the total plate-to-ground voltage e_a'. (The unprimed symbols e_v, e_a, and e_k will refer to incremental voltages with respect to ground rather than total voltages.) The plate and grid voltages listed on tube charts are always with respect to the cathode.

From eq. 9.7*b*, it can be seen that the plate-to-cathode voltage is made up of two voltage generators and a voltage drop through a resistance r_p. The equivalent of Fig. 9.3 can therefore be drawn directly from eq. 9.7*b*.

At higher frequencies, the small capacitances between the electrodes of the tube (as well as those in the leads, sockets, and wiring) become important, even though they may amount to only a few micromicro-

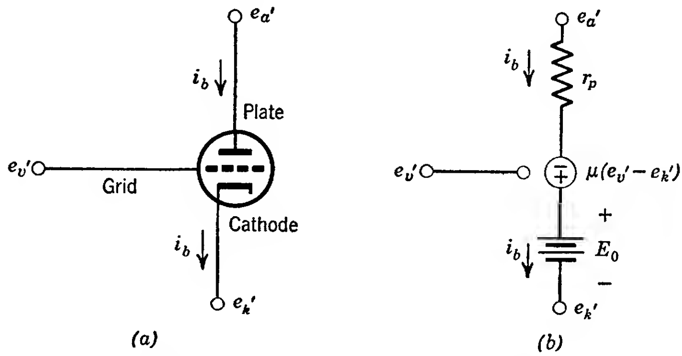


Fig. 9.3. Equivalent circuit of the triode.

farads. In addition, the inductances of the short wires connecting the electrodes of the tube to its socket become significant, especially the cathode lead inductance. A more complete equivalent circuit can be drawn by including the capacitances and the cathode lead inductance. The inductances of the plate and grid leads, although finite, are not usually included in the more expanded circuit shown in Fig. 9.4 because their effects are not particularly severe.

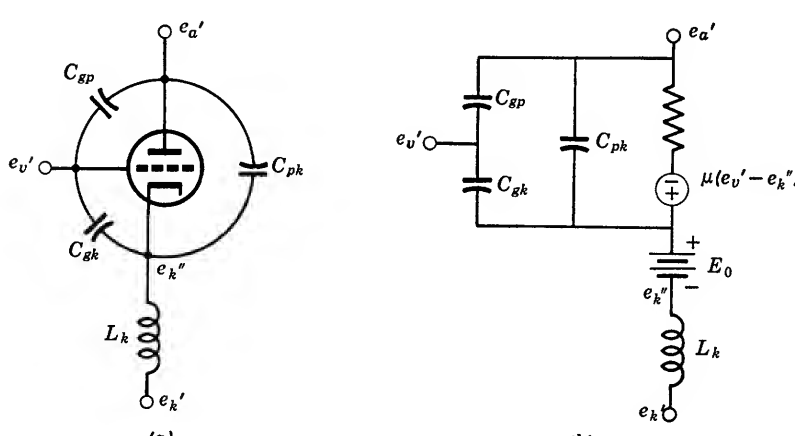


Fig. 9.4. A more exact triode equivalent circuit.

The true cathode voltage is the voltage within the envelope of the tube and not the tube pin terminal voltage which is separated from the true voltage by the inductance of the cathode lead wire.

If the voltage E_0 is ignored (as well as supply battery voltages), the resulting equivalent circuit is still valid except that average (d-c) levels are not specified. Average levels are of no interest in the sinusoidal case. In any event, they may be established using a strict d-c approach in which all direct voltages and currents are included and all frequency-sensitive circuit elements are entirely ignored, that is, inductances short-circuited (or replaced with an appropriate resistor to account for wire resistance) and capacitances removed. Generally, we shall ignore the d-c reference levels.

9.2 The triode at low frequencies

At low frequencies as well as at direct current, all stray and distributed capacitances and lead inductances can be ignored. The triode equivalent circuit then becomes quite simple. When the input to the triode is applied between grid and ground and the output is taken from

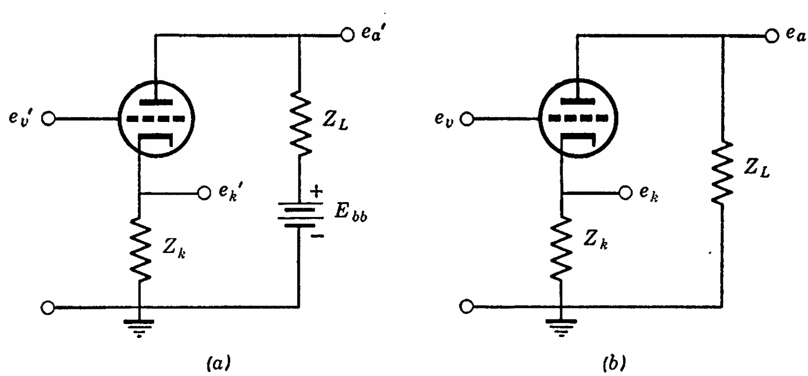


Fig. 9.5. Triode with plate and cathode impedances.

either the cathode or the plate (with respect to ground), the circuit diagram of Fig. 9.5*a* applies with Fig. 9.5*b* for alternating voltages where direct voltages are ignored. The ground (reference) node is assumed to be at a voltage of zero. The equivalent circuit of Fig. 9.5*b* is shown in Fig. 9.6*a*. Figure 9.6*b* is the same as Fig. 9.6*a* except that the voltage generator has been replaced with a current generator. The nodal equations for Fig. 9.6*b* are

$$\begin{aligned} (Y_L + 1/r_p)e_a - (1/r_p)e_k &= -g_m(e_v - e_k) \\ -(1/r_p)e_a + (Y_k + 1/r_p)e_k &= g_m(e_v - e_k) \end{aligned} \tag{9.8}$$

which become

$$
\begin{aligned}
(Y_L + 1/r_p)e_a - (g_m + 1/r_p)e_k &= -g_m e_v \\
-(1/r_p)e_a + (Y_k + g_m + 1/r_p)e_k &= g_m e_v
\end{aligned}
\tag{9.9}
$$

These equations characterize a feedback network. Had Z_k been zero, they would characterize an isolated network.

Equations 9.9 lead to the transfer functions

$$
\frac{e_k}{e_v} = \frac{g_m Y_L}{Y_L(Y_k + g_m + 1/r_p) + Y_k/r_p} = \frac{\mu Z_k}{(1+\mu)Z_k + r_p + Z_L} \tag{9.10}
$$

$$
\frac{e_a}{e_v} = \frac{-g_m Y_k}{Y_L(Y_k + g_m + 1/r_p) + Y_k/r_p} = \frac{-\mu Z_L}{(1+\mu)Z_k + r_p + Z_L} \tag{9.11}
$$

When the output is taken from the cathode, the circuit is called a "cathode follower." Normally, $Z_L = 0$ for cathode followers. Since μ

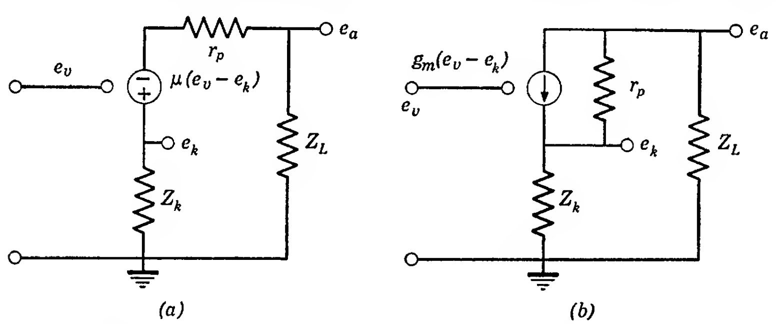


Fig. 9.6. Equivalent circuits of the triode with impedances.

is almost always fairly large (from 4 or 5 to perhaps 100), e_k/e_v is only slightly less than unity as long as Z_k is not too small. Should μ or r_p change somewhat with time or from one point in the applied a-c wave to another (nonlinear effects), the ratio e_k/e_v will change considerably less. The presence of the cathode impedance introduces negative feedback which provides this desirable behavior.

The input impedance (grid to ground) of the circuits of Fig. 9.6 is infinite because no impedances exist in the grid circuit. The output impedance (Thévenin circuit) will now be found for the cathode follower using a rather useful general method. If $Z_k = \infty$, the value of e_k/e_v is a maximum of $\mu/(1 + \mu)$. If Z_k is assumed to represent the total load

on the cathode follower, the special value of Z_k which causes e_k/e_v to be half its value at $Z_k = \infty$ defines the output impedance of the cathode follower Z_{out}; that is, when the source and load impedances of a circuit are equal, the output voltage is precisely one-half the source voltage. Thus eq. 9.10 yields

$$\frac{\mu}{2(1+\mu)} = \frac{\mu Z_{\text{out}}}{(1+\mu)Z_{\text{out}} + r_p + Z_L} \tag{9.12}$$

Solving for Z_{out}

$$Z_{\text{out}} = \frac{r_p + Z_L}{1+\mu} \tag{9.13}$$

which permits us to draw the equivalent circuit of Fig. 9.7 for the cathode follower where the impedance Z_k is the total load between

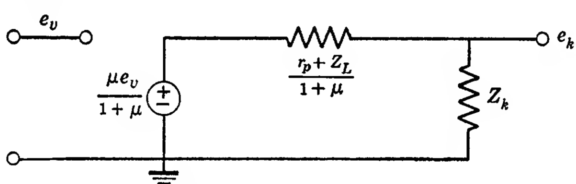


Fig. 9.7. Equivalent circuit for cathode output.

cathode and ground. It can be seen that the negative feedback provided by the cathode follower has reduced the output impedance by the factor $1/(1 + \mu)$ times the plate resistance of the tube. Although the voltage gain is somewhat less than unity, the power gain can be considerable because of the impedance level transformation from the high-impedance grid circuit to the low-impedance output circuit. Note also that the output impedance can be controlled with the plate circuit impedance Z_L.

If the output is taken from the plate, the voltage gain e_a/e_v can be large if Z_k is small; when $Z_k = 0$ and $Z_L \gg r_p$, it can be as large as μ.

The output impedance at the plate can be found as before. At $Z_L = \infty$, $e_a/e_v = -\mu$. At half this value, the output impedance Z_{out} is defined

$$-\frac{\mu}{2} = \frac{-\mu Z_{\text{out}}}{(1+\mu)Z_k + r_p + Z_{\text{out}}} \tag{9.14}$$

Solving for Z_{out}

$$Z_{\text{out}} = r_p + (1+\mu)Z_k \tag{9.15}$$

which is always larger than r_p and increases considerably for large Z_k. The equivalent plate output circuit is shown in Fig. 9.8.

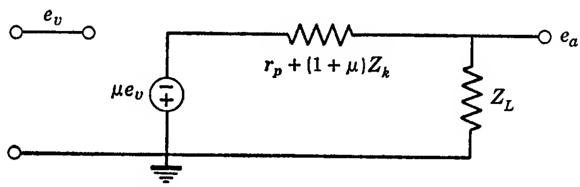


Fig. 9.8. Equivalent circuit for plate output.

A triode may be driven at its cathode with the output appearing at the plate and with the grid grounded as in Fig. 9.9. The arrangement

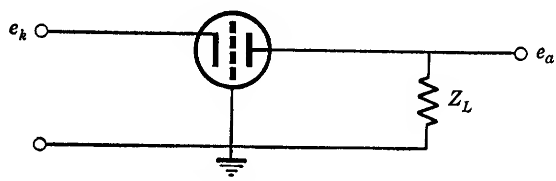


Fig. 9.9. The grounded-grid amplifier connection.

is then called a "grounded-grid" amplifier. The equivalent circuit of Fig. 9.9 is shown in Fig. 9.10. There is but a single node equation for

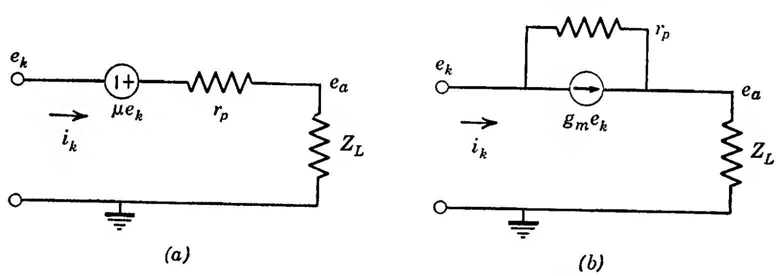


Fig. 9.10. Equivalent circuits for the grounded-grid amplifier.

this circuit (at the output) which leads directly to the transfer function as

$$\frac{e_a}{e_k} = \frac{(1 + \mu)Z_L}{r_p + Z_L} \tag{9.16}$$

At $Z_L = \infty$, $e_a/e_k = 1 + \mu$. As before, the output impedance is determined as

$$Z_{\text{out}} = r_p \tag{9.17}$$

which results in the equivalent circuit for the output shown in Fig. 9.11.

It should be noted that input and output signals for the grounded-grid circuit and the cathode follower have the same sign. The opposite sign applies when the input is at the grid and the output at the plate.

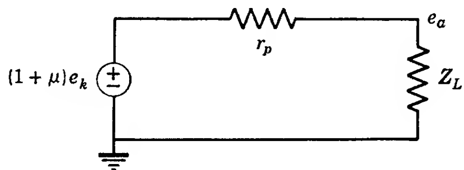


Fig. 9.11. Output circuit of the grounded-grid amplifier.

The grounded-grid circuit does not have an infinite input impedance. We may find the input impedance by obtaining the ratio of e_k/i_k as described in Fig. 9.10. The cathode current i_k is the sum of two currents easily obtained as

$$i_k = g_m e_k + (e_k - e_a)/r_p \tag{9.18}$$

Between eqs. 9.16 and 9.18, we obtain the desired ratio as

$$Z_{\text{in}} = \frac{e_k}{i_k} = \frac{r_p + Z_L}{1 + \mu} \tag{9.19}$$

which is usually a rather low input impedance compared to the plate resistance r_p and the load impedance Z_L. Between eqs. 9.17 and 9.19,

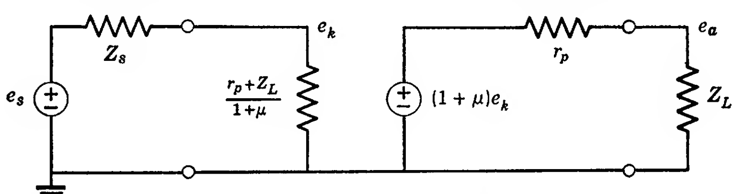


Fig. 9.12. Complete equivalent circuit of the grounded-grid amplifier.

we may draw a more complete equivalent circuit showing an external voltage source e_s as in Fig. 9.12 which drives the cathode of the tube. Z_s is the impedance of the external voltage source. This figure shows the character of both the input and output impedances of the grounded-grid circuit. For the special case when all impedances are matched, which requires $Z_L = r_p$ and $Z_s = (r_p + Z_L)/(1 + \mu)$, we get the overall transfer function from Fig. 9.12 as

$$\frac{e_a}{e_s} = \frac{1 + \mu}{4} \tag{9.20}$$

9.3 The effects of grid-to-cathode capacitance

At higher frequencies, stray and interelectrode capacitances cannot be ignored. Some of these capacitances can be accounted for fairly easily by including them in the circuits of the previous section. When the cathode is grounded and the output is taken from the plate, the plate-

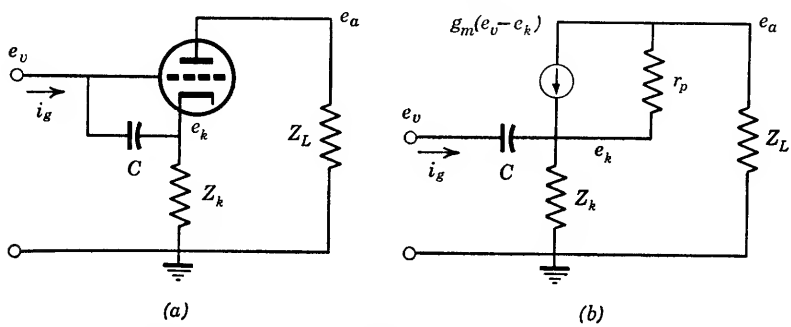


Fig. 9.13. Circuit with grid-to-cathode capacitance.

to-cathode and ground capacitances can be included as irreducible elements in shunt with the load impedance Z_L. More practically, load impedances can always be taken such that they include some capacitance in shunt. We shall study these specific capacitance effects again in Chap. 10 but are more interested in other phenomena here.

The grid-to-cathode capacitance has little effect on the circuit when the cathode is grounded (unless the cathode lead inductance is appreciable) but does provide an irreducible input impedance which is capacitive; that is, the input impedance to the grid cannot be assumed to be infinite. The grid-to-plate capacitance provides a feedback path from input to output, something we shall study specifically in the following section.

When the cathode is not grounded, some important effects which result from grid-to-cathode capacitance take place which warrant some study. In the interests of simplicity, all but the grid-to-cathode capacitance will be ignored in the analysis here.

Consider the circuit of Fig. 9.13 where C is the grid-to-cathode capacitance. The nodal equations are

$$\begin{aligned} (Y_L + 1/r_p)e_a - (g_m + 1/r_p)e_k &= -g_m e_v \\ -(1/r_p)e_a + (Y_k + g_m + pC + 1/r_p)e_k &= (g_m + pC)e_v \end{aligned} \tag{9.21}$$

From these equations, we get transfer functions as

$$\frac{e_k}{e_v} = \frac{Y_L(\mu + pr_pC) + pC}{Y_k + pC + Y_L(1 + \mu + Y_kr_p + pr_pC)} \tag{9.22}$$

$$\frac{e_a}{e_v} = \frac{-\mu Y_k + pC}{Y_k + pC + Y_L(1 + \mu + Y_kr_p + pr_pC)} \tag{9.23}$$

The current i_g can be found as

$$i_g = (e_v - e_k)pC \tag{9.24}$$

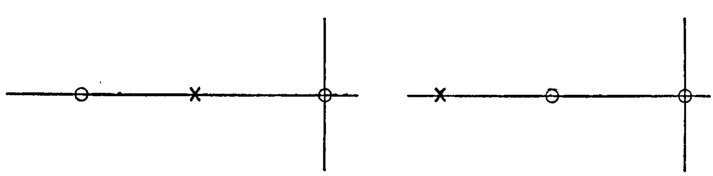
Between eqs. 9.22 and 9.24, the input admittance is found to be

$$Y_{\text{in}} = \frac{i_g}{e_v} = pC\left(1 - \frac{Y_L(\mu + pr_pC) + pC}{Y_k + pC + Y_L(1 + \mu + Y_kr_p + pr_pC)}\right) \tag{9.25}$$

It appears that when Z_k is not zero the input admittance can be reduced because of the negative sign in eq. 9.25. For the cathode follower where $Z_L = 0$ ($Y_L = \infty$) and assuming $Y_k = G_k + pC_k$ (parallel combination of capacitance and resistance), we get

$$Y_{\text{in}} = \frac{pCC_k}{C + C_k}\left\{\frac{p + [(1 + r_pG_k)/(r_pC_k)]}{p + \{(1 + \mu + r_pG_k)/[r_p(C + C_k)]\}}\right\} \tag{9.26}$$

which can be adjusted to show p-z as in Fig. 9.14*a* or *b*. The function of Fig. 9.14*b* cannot be realized with a passive *R-C* network because it



(a) (b)

Fig. 9.14. Input admittance with grid-to-cathode capacitance.

is capable of providing negative conductance (which can result in oscillations, particularly if inductance occurs in the grid circuit). The illustration should serve to indicate a method by which feedback can be utilized to obtain "unusual" input admittances.

If frequencies are low, the variable $p = j\omega$ will be small compared to the constant factors in eq. 9.26. Then, an approximation is

$$Y_{\text{in}} \cong pC\left(\frac{1 + r_pG_k}{1 + \mu + r_pG_k}\right) \cong \frac{pC}{1 + \mu} \tag{9.27}$$

where the second approximation is valid for small r_p/R_k. In this case, the equivalent input capacitance is only $1/(1 + \mu)$ times the actual grid-to-cathode capacitance. That this should occur is reasonable after all, because the cathode follower yields a cathode voltage that "follows" the grid voltage quite closely. Charging currents therefore cannot flow through the capacitance to any great extent; hence, the effective input capacitance cannot be large.

A grid resistor is often placed between grid and ground in order to control ion current and other extremely small second-order effects. If this resistor is returned from the grid to the cathode of a cathode follower rather than to the ground, an extremely high input resistance may be obtained at low frequencies, much higher than that of the physical grid-to-cathode resistor. This phenomenon is quite analogous to the reduction of input capacitance and is utilized in the construction of certain types of high impedance voltmeters.

At extreme frequencies where the variable p in eq. 9.26 is large compared to the constant factors, the input admittance becomes

$$Y_{\text{in}} \cong \frac{pCC_k}{C + C_k} \tag{9.28}$$

which is a capacitance given by the series combination of C and C_k as might be expected; the gain of the cathode follower becomes so small that the full charging current can flow.

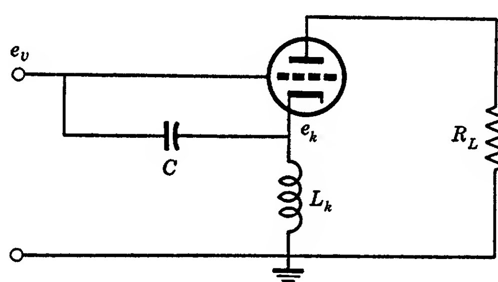


Fig. 9.15. Triode with cathode lead inductance.

If Z_L is not zero, the input admittance will also depend upon Z_L. The major effect is that the input capacitance is not reduced as much at low frequencies as when $Z_L = 0$.

When the cathode lead inductance cannot be neglected, a more insidious effect, partly dependent upon the grid-to-cathode capacitance, takes place. In most tubes it is this effect that limits the highest fre-

quencies at which amplification can be obtained without resorting to special measures.

The impedance Z_k of previous equations can be assumed to consist of the cathode lead inductance as shown in the circuit of Fig. 9.15. By setting $Z_k = pL_k$ and $Y_L = G_L$ (a purely resistive plate load) in eq. 9.25, we obtain

$$Y_{\text{in}} = \frac{G_L}{1 + r_pG_L}\left\{\frac{p[p + (1 + r_pG_L)/(L_kG_L)]}{p^2 + \{[(1 + \mu)G_L]/[(1 + r_pG_L)C]\}p + [1/(L_kC)]}\right\} \tag{9.29}$$

which has a p-z plot as shown in Fig. 9.16. It is possible for the real part of this input admittance to become very large, thus furnishing a relatively low input resistance. Replacing p by $j\omega$ in eq. 9.29 and solving for the real part, we get

$$G_{\text{in}} = \text{Re } Y_{\text{in}}(j\omega)$$

$$= \frac{1}{r_p + R_L}\left\{\frac{\omega^2(\omega^2 + \mu\omega_0^2)}{\omega^4 + \omega^2\{(1 + \mu)^2/[(R_L + r_p)^2C^2] - 2\omega_0^2\} + \omega_0^4}\right\} \tag{9.30}$$

where $\omega_0 = (1/L_kC)^{1/2}$ is the series-resonant frequency of L_k and C. For small ω, eq. 9.30 is approximately

$$G_{\text{in}} \cong g_m\left(\frac{r_p}{r_p + R_L}\right)\left(\frac{\omega}{\omega_0}\right)^2 \tag{9.31}$$

where the approximation neglects the term ω^4 in the numerator and all but ω_0^4 in the denominator. The conductance G_{in} of eq. 9.31 grows as the square of frequency. Since $1/g_m$ often represents a relatively small resistance (perhaps as low as 50 or 100 Ω), the reduction of the input impedance resulting from cathode lead inductance can be serious.

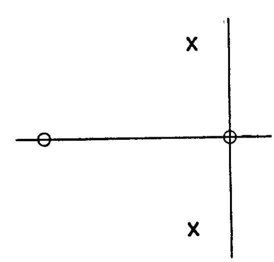


Fig. 9.16. Input admittance due to cathode lead inductance.

Certain techniques minimize this effect. A small external capacitance (or capacitive transmission-line segment) can be placed between the cathode tube pin and ground so that the series-resonant circuit thus formed has zero impedance or a small resistance at the frequency of interest. The result is that the cathode-to-ground impedance becomes very small at

the series-resonant frequency so that the input impedance at the grid has only a reactive component with perhaps only a residual conductance component. This method of by-passing the cathode to ground is referred to as "resonant" by-passing and might appear in a circuit as in Fig. 9.17.

It is possible to actually reduce the magnitude of the cathode lead inductance by bringing out heavy cathode leads at many points and/or making the cathode lead a continuous sheet designed to be part of a resonant circuit (as is done in "disc-seal" tubes).

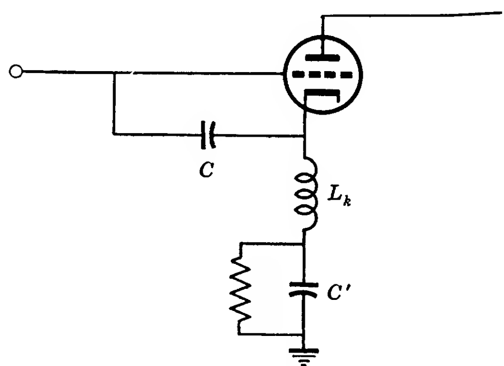


Fig. 9.17. Resonant by-passing.

A different phenomenon having an effect like that resulting from cathode lead inductance is transit time. When the electrons require a finite time to travel from cathode to grid and from grid to plate relative to the periods of the sine-wave voltages and currents applied to the tube, another input conductance component is introduced which increases as the square of frequency. This component can be minimized only by operating the tube at higher voltages or by using physically smaller tubes. It is this component that sets the ultimate upper frequency limit of a tube. Since the electron transit time effect gives a conductance increasing as the square of frequency, it can be accounted for as far as its effect on the input impedance at the grid is concerned by assuming an additional component of cathode lead inductance that cannot be made impotent by means of resonant by-passing. However, an appreciable transit time also has a severe effect on the gain of a tube (the effective g_m and μ), reducing it rapidly at frequencies where the transit time is comparable to the periods of the alternating voltages in the tube. (Not only is the effective gain reduced by transit time but a time lag is introduced; that is, a negative phase shift must be ascribed to the transconductance.)

The effect of cathode lead inductance on the grounded-grid circuit is comparable to that already discussed. However, it is not so noticeable (relatively) because the grounded-grid system operates at such a low impedance level. In addition, the cathode lead inductance can more easily be made an inherent part of the circuit driving the cathode. Triodes are often operated as grounded-grid amplifiers rather than as grounded-cathode amplifiers at high frequencies because problems brought about by the cathode lead inductance are reduced (although another important objective is to avoid the Miller effect).

9.4 The Miller effect

When the grid-to-plate capacitance of a grounded-cathode stage is appreciable, a special kind of feedback exists from which results the so-called Miller effect. Specifically, it is the equation giving the input impedance at the grid of a tube having a finite grid-to-plate capacitance that is termed the Miller effect.

"Neutralization" can be employed to reduce the effect by shunting the grid-to-plate capacitance with an inductance so that the parallel combination is resonant at the frequency of interest. Then, a very high grid-to-plate impedance is given at that frequency, thereby minimizing the feedback.

Because the Miller effect is so pronounced in triode tubes, high-frequency grounded-cathode amplifiers rarely incorporate triodes unless

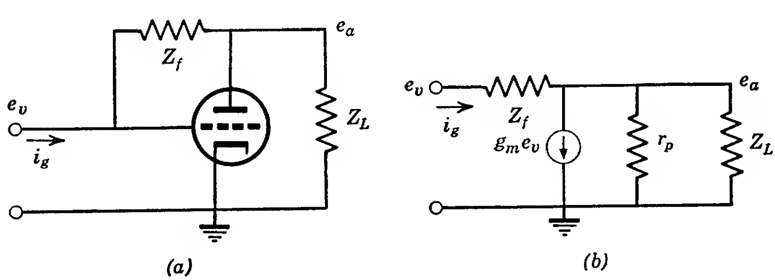


Fig. 9.18. Generalized circuit with grid-to-plate impedance.

neutralized. As grounded-grid amplifiers, either triodes or pentodes may be used (with triodes best where noise is a limiting factor). In a grounded-grid triode, the grid acts to shield the cathode and plate so that there is but a small capacitance between input and output (and the input and output voltages are in phase rather than 180 degrees out of phase).

In many cases, the Miller effect is desirable and makes possible useful

circuits. For example, it can be employed to obtain negative capacitance, large positive capacitance, or other "unusual" circuit impedances. When generalized, it becomes the essential phenomenon in analog computer amplifiers.

Consider the circuit of Fig. 9.18, which is that of a grounded-cathode amplifier with generalized grid-to-plate and plate impedances. The one node equation that is required leads to

$$\frac{e_a}{e_v} = \frac{-\mu + r_p Y_f}{1 + r_p(Y_L + Y_f)} \tag{9.32}$$

As before, the output impedance can be obtained as

$$Z_{\text{out}} = \frac{r_p}{1 + r_p Y_f} \tag{9.33}$$

The current i_g is given by

$$i_g = (e_v - e_a)Y_f \tag{9.34}$$

Between eqs. 9.32, 9.33, and 9.34, the input admittance can be determined as

$$Y_{\text{in}} = Y_f\left(\frac{1 + \mu + r_p Y_L}{1 + r_p(Y_L + Y_f)}\right) \tag{9.35}$$

When the circuit of Fig. 9.18 is connected to a source of voltage e_s with a source impedance Z_s, the complete equivalent circuit becomes that shown in Fig. 9.19.

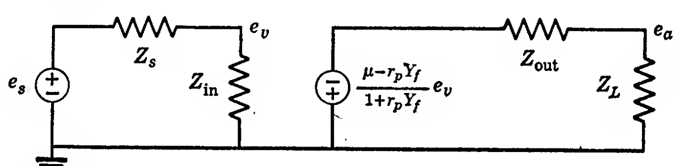


Fig. 9.19. Complete equivalent circuit with plate-to-grid feedback.

To study the most common type of Miller effect, assume $Y_f = pC_f$ (capacitance) and $Y_L = G_L + pC_L$ (capacitance and conductance in parallel). Then eq. 9.35 becomes

$$Y_{\text{in}} = \frac{C_L C_f}{C_L + C_f}\left\{\frac{p\{p + [(1 + \mu + r_p G_L)/(r_p C_L)]\}}{p + \{(1 + r_p G_L)/[r_p(C_L + C_f)]\}}\right\} \tag{9.36}$$

which has a p-z diagram as shown in Fig. 9.20.

At high frequencies, Y_{in} becomes the series combination of C_f and C_L as might be expected. At low frequencies, p in the quantity in braces in eq. 9.36 can be ignored and we get

$$Y_{\text{in}} \cong pC_f \frac{1 + \mu + r_pG_L}{1 + r_pG_L} \tag{9.37}$$

which is a capacitance that (for r_pG_L small) is $1 + \mu$ times as large as the physical grid-to-plate capacitance C_f. Such a high input capacitance causes the gain to become very small as frequencies are increased to even modest values. Approximately, the input circuit of Fig. 9.19 acts like the low-pass filter shown in Fig. 9.21. If $Z_s = R_s$, the gain of this circuit falls to $1/(2)^{1/2}$ of that at very low frequencies at a frequency $\omega = 1/(R_sC)$.

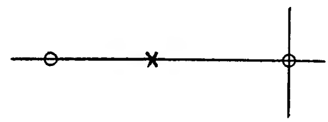


Fig. 9.20. Input admittance with plate-to-grid feedback.

Although such a high capacitance may be objectionable in some cases, it may sometimes be useful because it provides a means for obtaining much larger capacitances than would be practical with bilateral circuit elements. In addition, the capacitance can be made variable by providing suitable means for changing the amplification of the tube.

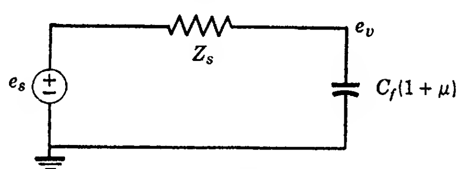


Fig. 9.21. Input circuit with grid-to-plate capacitance.

A cascade of two grounded-cathode tubes can be employed to make the gain of the combination at low frequencies effectively positive. Then, if a capacitor is placed between input and output of the combination, the input capacitance can be caused to become negative. This technique can be employed to cancel out certain objectionable capacitances; however, it must be remembered that as frequencies are raised the ability to cancel capacitance in this manner is reduced because of the dropping off of gain resulting from other distributed and interelectrode capacitances.

The equivalent circuit of a tube with generalized Miller effect and source and load impedances is shown in Fig. 9.22. From the two node equations required for the solution to this circuit, we can easily find the

transfer function as

$$\frac{e_a}{e_s} = -\frac{Y_s/Y_f}{1 + \dfrac{Y_f + Y_s}{Y_f}\left(\dfrac{1 + r_p(Y_L + Y_f)}{\mu - r_pY_f}\right)} \cong -\frac{Y_s}{Y_f} \tag{9.38}$$

where the approximation is valid for large μ at low frequencies. The final result shows the behavior to depend only on the admittances Y_s/Y_f

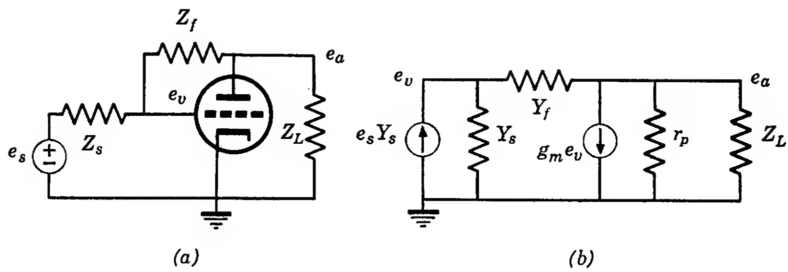


Fig. 9.22. Miller circuit with external voltage source.

and is the basis for electric analog computing amplifiers; negative feedback is carried to such an extreme that the over-all system behavior is essentially independent of moderate nonlinearities and instabilities associated with the vacuum tubes in the system.

9.5 Pentode equivalent circuits

A tetrode is essentially a triode with an additional grid between the control grid and the plate called the screen grid. At certain electrode voltages, electrons hitting the plate of a tetrode can be returned to the cathode region along with extra electrons "shaken off" the plate, which is normally considered to be an undesirable effect. By placing a third grid between the plate and the screen grid (the suppressor grid) and operating this grid at the potential of the cathode, the "secondary" electrons shaken off the plate meet a retarding electric field and are ultimately returned to the plate. Except for certain electrode potentials, pentodes (with three grids) and tetrodes (with two grids) have quite similar characteristics. Normally, the screen grid is operated at a potential on the same order of magnitude as that of the plate, perhaps a little less.

The presence of grids between the control grid and the plate yields two very important advantages in grounded-cathode operation. First, the screening effect reduces the control-grid-to-plate capacitance to a very small value, which greatly reduces unintentional Miller effects

(although the grid-to-plate capacitance may still be large enough to introduce problems in high-frequency systems). Second, when the screen grid is operated at a constant potential, changes in the plate voltage have very little effect on the plate current. Then, the lines of constant grid voltage will appear almost horizontal on the plate-voltage plate-current characteristics.

Normally, the suppressor grid is operated at ground or cathode potential and frequently is connected to the cathode within the envelope of the tube. Although it may sometimes be used as an on-off control elec-

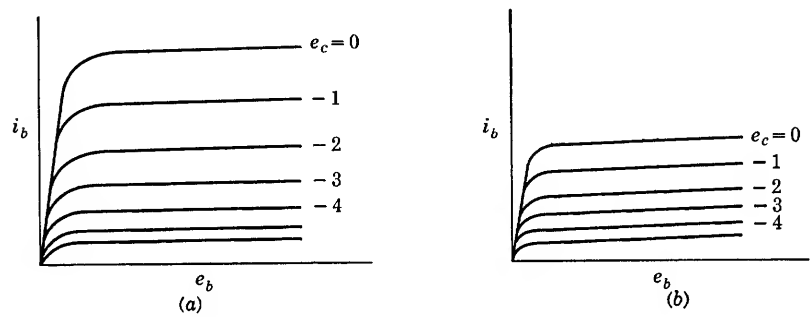


Fig. 9.23. Pentode plate characteristics, where e_{c2} is large (*a*) or small (*b*).

trode, we shall assume that it does not play a part in the operating behavior of the pentode.

The screen grid will intercept some of the electrons in such a way that the plate current in a pentode is less than the cathode current. In fact, the control grid and screen grid act somewhat like a triode.

The plate current in a pentode is dependent upon the screen-grid voltage as well as on the control-grid and plate voltages. This introduces an additional variable to be accounted for in the operation of the pentode as compared to the triode. When the voltage at the screen is constant, we may use a set of characteristic curves similar to those for the triode, a representative set of which is shown in Fig. 9.23*a* for one particular value of screen voltage. However, a different set of curves must be used for a different screen voltage, as is indicated in Fig. 9.23*b*.

The screen grid does draw current and it may be necessary to know the screen current as a function of the grid voltage. For example, if an impedance exists in the screen circuit, the screen current and voltage can change with grid voltage with the result that operation is not confined to one set of characteristic curves.

The linearized curves for the pentode take the form of a number of parallel and equidistant lines as is the case for the triode. Such a set of

curves appears in Fig. 9.24. For the linear approximation to be at all valid, it must be assumed that operation is confined to the right of the zero grid-voltage line of this figure.

When the screen voltage is constant, only one set of characteristic curves applies and the equations of behavior are quite similar to those

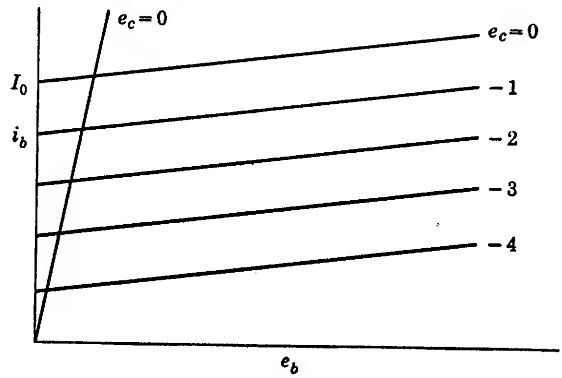


Fig. 9.24. Linearized pentode characteristics.

for the triode, that is

$$i_b = C_0 + \frac{\partial i_b}{\partial e_c} e_c + \frac{\partial i_b}{\partial e_b} e_b = C_0 + g_m e_c + \frac{e_b}{r_p} \tag{9.39}$$

where all voltages are with respect to the cathode. The constant C_0 can be evaluated by observing that, at $e_b = e_c = 0$, $i_b = I_0$. Thus $C_0 = I_0$.

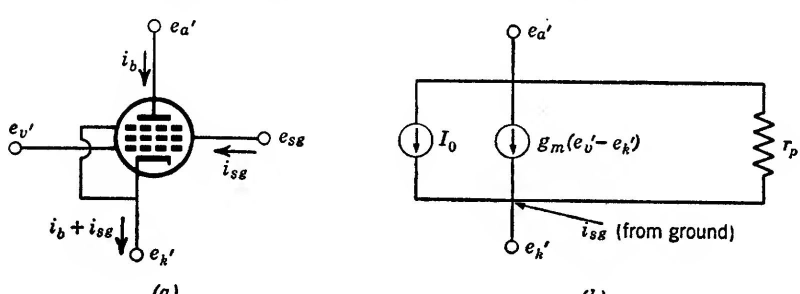


Fig. 9.25. Equivalent circuit of the pentode.

It is more convenient to use a current generator equivalent circuit for pentodes. The equivalent circuit of Fig. 9.25 can be deduced from eq. 9.39. In this circuit I_0 is constant and can be neglected when only time-varying quantities are involved. However, i_{sg} is not always con-

stant with e_v'. As a result, a secondary effect will be observed if there is an impedance from cathode to ground. Fortunately, i_{sg} is usually small enough to be neglected.

The circuit resulting when I_0 and i_{sg} are neglected is identical to the triode circuit already described and all the considerations relating to circuitry studied in the previous sections are directly applicable. The only differences are in the typical values of r_p, μ, and g_m (and the grid-to-plate capacitance). The transconductance g_m of triodes and pentodes is about the same, perhaps a few hundred to 10,000 or 20,000 micromhos. The plate resistance of triodes r_p ranges from a few hundred ohms to perhaps 100K, whereas that for pentodes is much larger, from a few hundreds of thousands to a few millions of ohms. Since $\mu = g_m r_p$, the amplification factor of pentodes is considerably larger than that of triodes.

The plate resistance of a pentode is so large that it is frequently neglected in comparison with other circuit elements that may be in parallel with it, as is almost always done in connection with the study of wide-band amplifier circuits (which always employ relatively small plate load impedances).

When the screen voltage is not constant and/or the current i_{sg} is not near enough constant or small enough to be ignored, suitable modifications must be employed. In eq. 9.39 an additional term

$$\frac{\partial i_b}{\partial e_{c2}} e_{c2} \tag{9.40}$$

must be included. This adds another current generator to the equivalent circuit. When an impedance exists in the screen circuit, the voltage e_{c2} will be dependent upon i_{sg}. Consequently, an additional linear equation is required to describe the screen current. It has the form

$$i_{sg} = C_0' + \frac{\partial i_{sg}}{\partial e_c} e_c + \frac{\partial i_{sg}}{\partial e_b} e_b + \frac{\partial i_{sg}}{\partial e_{c2}} e_{c2} \tag{9.41}$$

The simultaneous solution of the equations for i_b and i_{sg} leads to the more complete equivalent circuit. The results are not at all simple and the solution is useful only for quite special cases. Rather than derive the relations, some approximate results will be set down. When an impedance Z_{sg} exists between the screen grid and ground, the transconductance g_m (referring to the plate) can be replaced with another transconductance

$$g_m \longrightarrow g_m \frac{r_s}{r_s + Z_{sg}} \tag{9.42}$$

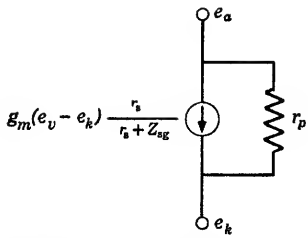


Fig. 9.26. Equivalent circuit including a-c screen effects.

where r_s is the source resistance of the screen, which is analogous to the plate resistance of a triode and is typically on the order of 10K or 20K. A more complete equivalent circuit can now be drawn for the pentode (a-c variations only) as shown in Fig. 9.26.

It should be pointed out that the screen grid can be used as a control grid; that is, an input signal can be applied to the screen grid rather than to the control grid. In fact, both grids can be utilized so that two signals may be applied simultaneously.

9.6 Certain combinations of tubes

In many instances, desirable circuits may be obtained by operating two or more tubes in some parallel arrangement. For example, two triodes can be made to add two signals as shown in Fig. 9.27. A sub-

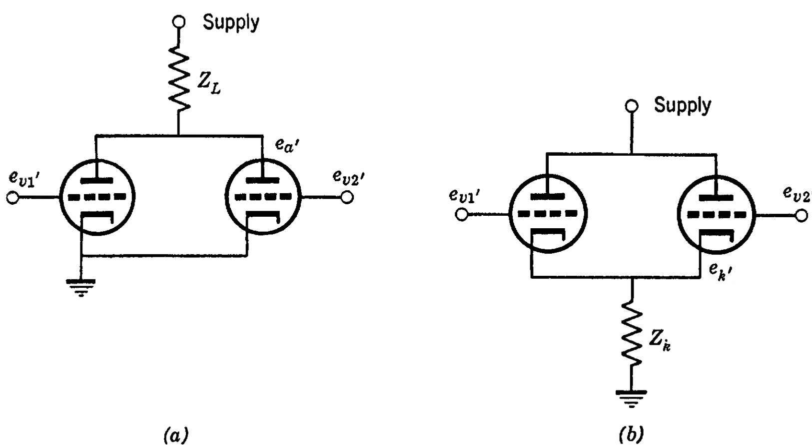


Fig. 9.27. Two-tube adding circuits.

traction circuit might have the form shown in Fig. 9.28, which also provides a means of obtaining high gain with a high impedance input and without a sign reversal from input to output. The circuit of Fig. 9.28 (or some minor modification of it) is widely used as an input system to d-c amplifiers because fluctuations in supply voltages and currents and in tube characteristics tend to be cancelled between the two tubes.

When an input signal exists on a pair of wires such that the system is balanced with respect to ground, tubes may be operated in a "push-

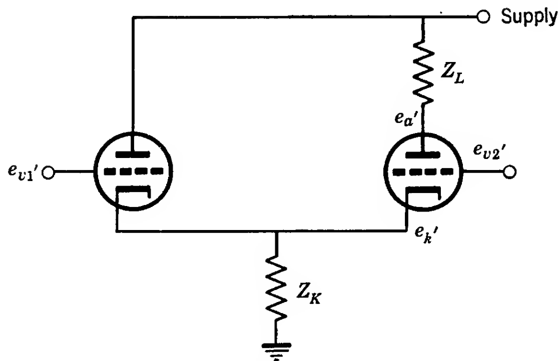


Fig. 9.28. A two-tube subtracting circuit.

pull" arrangement as indicated in Fig. 9.29. If $Z = Z'$ in this circuit, the system behaves like two separate systems with respect to ground with equal and opposite voltages applied to the two inputs. Fluctuations in tube characteristics and supply voltages tend to cancel in a push-pull system. Push-pull systems not having balanced circuitry enable networks such as bridges to be realized.

In all multiple-tube circuits such as those mentioned, equivalent circuits permit an analysis to be made in a straightforward manner. Particular attention should be paid to the d-c characteristics to insure that the tubes will work together harmoniously; that is, one tube should not force undesirable direct voltages and currents on the other, and so forth.

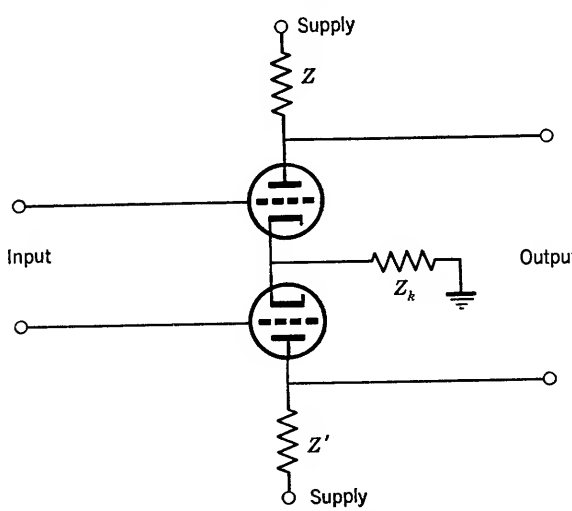


Fig. 9.29. The push-pull circuit.

9.7 Setting the operating point

When the potentials and currents about which the tube is to be varied with a signal are determined, a strict d-c approach is called for in which all inductances are considered short circuits (or resistors) and all capacitors open circuits. If the tube is a triode, the d-c circuit usually has the form of Fig. 9.30, where E_{bb} is the power supply voltage and E_{cc} is

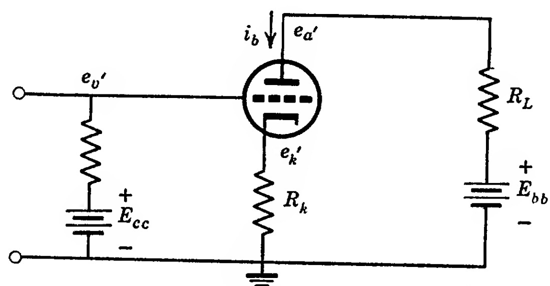


Fig. 9.30. A tube with generalized biasing.

a direct voltage applied by some means between grid and ground. Usually, E_{cc} is zero, although it may occasionally be made positive by attaching the grid to a voltage divider between E_{bb} and ground. The operating point is obtained from a graphical analysis on the tube-characteristic curves.

The product $(e_a' - e_k')i_b = e_b i_b$ is the d-c power dissipated at the plate of the triode. It is not safe to exceed a certain value of power

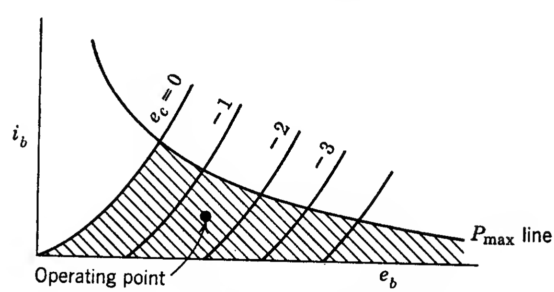


Fig. 9.31. Maximum-dissipation line.

which is called the "maximum plate dissipation." The curve of maximum permissible power $P_{max} = (e_a' - e_k')i_b$, where P_{max} is a constant, can be constructed on the characteristic curves as indicated in Fig. 9.31. The operating point is restricted to the shaded area of this figure, which is to the right of the $e_c = 0$ line in the interests of linearity, and

below the P_{max} line so that the safe ratings of the tube will not be exceeded. (It should be remembered that the abscissa of all tube-characteristic curves are in units of voltage with respect to the cathode.)

The operating point should be chosen somewhere in the middle of the permissible region of the characteristics. It is not good practice to operate a tube too close to its maximum rating, nor should operation be too near the zero bias line unless the expected magnitudes of the sine-wave voltages and currents in the tube are quite small. If the operating point occurs at a very large value of e_b, an excessively large power supply voltage will be required.

Two equations describing Fig. 9.30 are readily seen to be

$$\begin{aligned} E_{bb} - i_b R_L &= e_a' \\ i_b R_k &= e_k' \end{aligned} \tag{9.43}$$

Since the abscissa of the characteristic curves is expressed in units of $e_a' - e_k'$ and the grid bias in units of $e_c = E_{cc} - e_k'$, eqs. 9.43 can advantageously be rephrased as

$$E_{bb} - i_b R_L - e_k' = E_{bb} - i_b(R_L + R_k) = e_a' - e_k' = e_b \tag{9.44a}$$

$$i_b R_k = E_{cc} - (E_{cc} - e_k') = E_{cc} - e_c \tag{9.44b}$$

where e_b and e_c are values with respect to the cathode.

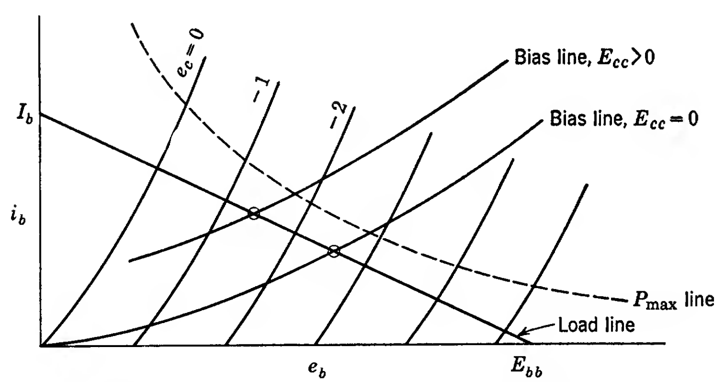


Fig. 9.32. Graphical solution for the operating point.

The curve given by eq. 9.44a can be plotted as a straight line on the characteristic curves. This curve is called the "load line." At $i_b = 0$, $e_b = E_{bb}$, whereas at $e_b = 0$, $i_b = E_{bb}/(R_L + R_k) = I_b$. These two points define the load line as is indicated in Fig. 9.32. *The operating point must lie somewhere on this line.*

Equation 9.44*b* may also be plotted on the characteristic curves. To plot this curve, a value of e_c is assumed and i_b is calculated to give a point. (It is assumed that E_{cc} is constant and is known.) This procedure is repeated for a sufficient number of values of e_c to define a curve. Such a curve, which can be called the "bias line," is also shown in Fig. 9.32. *The operating point must lie somewhere along this line.*

Since eqs. 9.44*a* and *b* must be satisfied simultaneously, the operating point must occur at the intersection of the two curves.

When a resistor occurs in the cathode in the manner described, the tube is said to have "cathode bias" (at least in part). If $R_k = 0$, e_a' becomes e_b and the load line is drawn the same as before. A bias line is then not required. However, a negative voltage with respect to the cathode must somehow be applied to the grid. This voltage is the "fixed bias" e_c shown on the characteristic curves. To obtain a fixed bias, a battery can be placed in the grid circuit, the grid resistor can be returned to a negative supply voltage, or the cathode can be operated at a fixed positive voltage with respect to the grid. The analysis of the operating point of the cathode follower is done as described; both the load line and the bias line are required.

If the d-c resistance of the plate load impedance is zero, as when the load is a parallel-resonant circuit, then the load line for d-c considerations will be determined only by the cathode resistor R_k. If R_k is also zero, the load line will be vertical. For steady-state sine-wave operation, the plate load impedance may be finite even though the d-c impedance is zero. Then the load line for a-c variations will pass through the operating point with a slope determined by the magnitude of the a-c impedance. (Strictly speaking, the load line will be a "line" only if the a-c impedance is purely resistive. Otherwise, it will be a more complicated two-dimensional curve such as an ellipse.)

The operating point of a pentode is obtained in a quite analogous manner if it is assumed that the screen voltage is constant at the value specified on the set of characteristic curves. The only slight modification required (which is often ignored) is due to the screen current, which causes the cathode current to be slightly larger than the plate current.

The screen is generally supplied in one of three different ways, which are shown in Fig. 9.33. The method shown in Fig. 9.33*a* may be the most costly way of supplying the screen unless power supplies are already available. The method of Fig. 9.33*b* is not good unless the plate supply voltage is low enough to be within the maximum rating of the screen; however, it is the simplest and cheapest way for providing the screen with voltage. The method of Fig. 9.33*c* is a common one. The value

of the screen dropping resistor is calculated by requiring it to drop the voltage from the supply voltage to the desired screen voltage while carrying the proper screen current.

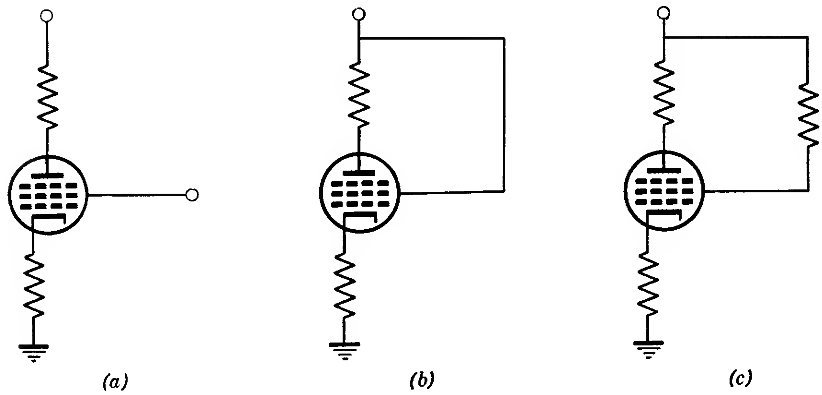


Fig. 9.33. Various methods for supplying the screen. (a) Separate screen supply. (b) Common plate and screen supply. (c) Screen dropping resistor.

Another method for obtaining grid bias should be mentioned for completeness and because it is quite common in some types of amplifying systems. It would appear at first glance that the circuit of Fig. 9.34 provides no means for holding the grid at an average negative potential. This will indeed be the case when the source voltage e is zero. However, when e is a sine wave, the grid will be positive with respect to the

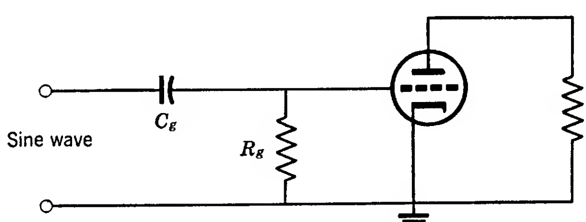


Fig. 9.34. Triode with signal bias.

cathode during the positive peaks of the applied sine wave and grid current will flow. In fact, the grid acts like the plate of a diode tube. This will leave a negative charge on the capacitor C_g which slowly leaks off through the "grid-leak" resistor R_g. Then the grid will have an average negative potential. The charge that leaks off C_g is replenished at each positive peak of the applied sine wave. As long as the amount

of grid current that flows at the peak of each cycle is small, the nonlinear distortion that is introduced will not be serious.

The biasing method described in the foregoing is referred to as "signal bias" or "grid-leak bias" and depends on the presence of an applied signal voltage. It is sometimes used in addition to a small amount of cathode or fixed bias, particularly if the loss of signal would result in excessive current in the tube.

Sometimes in low-cost audio voltage amplifiers a circuit like that of Fig. 9.34 will be found in which R_g is on the order of 10 megohms. In such circuits, biasing is not obtained from the signal; rather, the minute second-order electron and ion currents flowing in the very large grid-leak resistor are sufficient to provide a moderate negative voltage on the grid.

9.8 Peak voltage and power

Once the operating point has been located, the peak values of voltage and current obtainable at the output of the tube without excessive non-linearity may, for the case of resistive loading, be calculated.

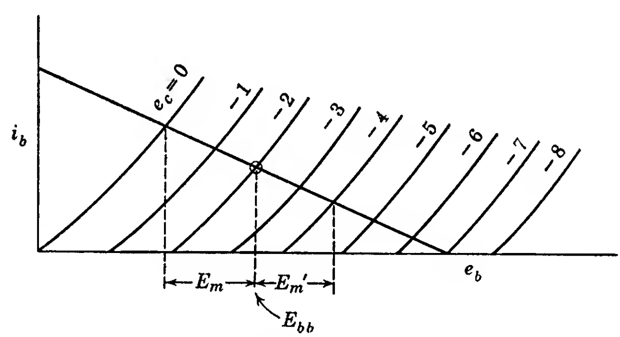


Fig. 9.35. Large signal limits for linearity.

Consider Fig. 9.35, which shows a load line and the operating point on a set of triode characteristics. It is assumed that the output is taken from the plate. If a sinusoidal voltage is applied to the grid of the tube such that e_c is varied between maximum limits of 0 and -4 volts (for example), the output at the plate will vary from $E_{bb} - E_m$ to $E_{bb} + E_m'$. If E_m and E_m' are not equal, the output sine wave will be somewhat distorted. If E_m and E_m' are equal and the separation of the grid curves is constant, there will be negligible distortion and the sine-wave power dissipated in the load resistance R_L will be $E_m^2/2R_L$. This power cannot be materially increased without driving the grid harder, with the result that e_c becomes positive on one peak of the sine wave and/or e_c becomes

so negative on the other peak that the plate current is reduced to zero. This latter condition is specified in terms of the "cutoff" bias, which is -7 volts for the example of Fig. 9.35. For positive e_c, grid current will flow in the tube, which results in one peak of the applied sine wave of voltage being more or less flattened. If e_c becomes more negative than -7 volts, the other peak of the sine wave will be rather abruptly flattened. Thus it is important that excursions of e_c be limited between zero and the cutoff bias if severe nonlinear behavior is to be avoided.

Operation of a tube in the manner described in Fig. 9.35 and in the foregoing discussion is termed "class-A" operation. It is of interest to

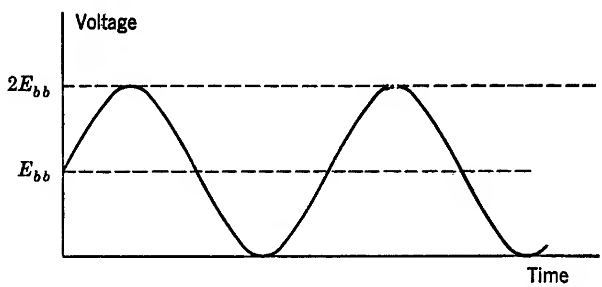


Fig. 9.36. Maximum signal in class-A amplifiers.

calculate the maximum possible conversion efficiency of the tube for this mode of operation. At the very best, E_m can be made as large as E_{bb}, which requires among other things that $R_k = 0$. The waveform across R_L would then appear as in Fig. 9.36. The d-c power supplied to the tube is E_{bb}^2/R_L if the load resistor is driven with a transformer or in parallel with an inductor such that the operating point occurs at a plate voltage equal to the power supply voltage. (If the plate current is supplied through the load resistor, the d-c power will be $4E_{bb}^2/R_L$, the required power-supply voltage being $2E_{bb}$.) The a-c power in the sine wave of Fig. 9.36 is $E_{bb}^2/2R_L$. Hence the maximum possible conversion efficiency is the sine-wave power divided by the d-c power, or 50 per cent. (It is only 25 per cent if the plate current is supplied through the load resistor.) The maximum possible efficiency of class-A pentodes is slightly less than this because of losses at the screen. For any applied voltage less than that giving the maximum possible swing just discussed, less efficient operation of course results. Efficiencies of class-A power amplifiers designed to handle most signals are rarely made greater than about 20 per cent in practice because of the distortion introduced. With feedback, efficiencies of around 40 per cent are practical. When a class-A power amplifier is used with pulse-type signals,

efficiencies near 50 per cent may be obtained. If the tube is operated as an on-off device rather than as a linear amplifier, efficiencies above 50 per cent may be obtained.

The maximum output voltage of the cathode follower cannot be larger than that required to cause cutoff if distortion is to be avoided, as occurs when the plate current goes to zero. If I_b is the current at the operating point and R_k is the cathode resistor, I_bR_k is the maximum peak output voltage that can be achieved, which corresponds to a sine-wave power of $I_b^2R_k/2$.

A sine wave causes the instantaneous values of i_b and e_b to move along some path on the characteristic curves. If the sinusoidal varia-

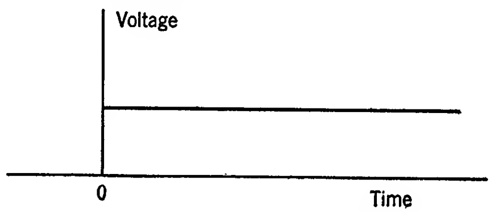


Fig. 9.37. Step-voltage input.

tions are large, the various tube constants, r_p and g_m (and to a lesser extent μ), vary as well. This causes some distortion to be introduced because the vacuum-tube circuit changes from point to point of the applied waveform. Numerous effects are introduced which we do not have time to study. However, one example will be indicated in order to point out what may take place.

Consider the application of a moderately large change of voltage to the grid of a cathode follower at time $t = 0$ as depicted in Fig. 9.37. If this change is positive, the tube will draw more current after $t = 0$ with the result that r_p will be decreased and g_m increased; consequently, the output impedance of the tube will be decreased. On the other hand, if this change is negative after $t = 0$, r_p will be increased, g_m decreased, and the output impedance will be increased, perhaps drastically.

Problems

1. Determine the transfer function e_a/e_v for the circuit of Fig. 9.5*b* when $Y_L = G + pC$ and $Y_k = G_k + pC_k$ and plot the p-z. Determine the maximum value of the transfer function. Also determine the maximum value when $RC \ll R_kC_k$.

2. Determine the transfer function e_a/e_v for the circuit of Fig. 9.5*b* when $Y_L = G + pC + 1/pL$ and $Y_k = G_k + pC_k + 1/pL_k$ and plot its p-z. Find the magnitude of the transfer function at resonance and specify the half-power bandwidth.

3. A cathode follower is used to drive a 50 Ω load. If $g_m = 5000$ micromhos and $r_p = 10\text{K}$, determine the voltage gain e_k/e_v and also the power gain assuming the impedance of the grid circuit is 500K.

4. If R_k for a cathode follower is returned to ground as is also the grid resistor, what limits the upper value of R_k? Can you suggest two ways to increase R_k (by means of various d-c connections) without reducing the transconductance of the tube?

5. A grounded-grid stage has a load admittance $G + pC + 1/pL$. Determine the expression for the input admittance and plot its p-z. Compare the bandwidth of the input admittance to that of the plate load.

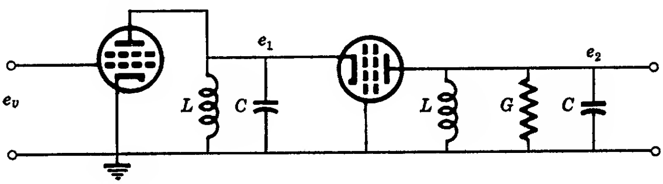


Fig. P.6.

6. The "cascode" amplifier is shown in Fig. P.6. It consists of a grounded-cathode stage followed by a grounded-grid stage. The input impedance of the grounded-grid stage is usually low enough to permit assuming the bandwidth of the plate load of the grounded-cathode stage to be much larger than that of the grounded-grid stage. Determine the over-all bandwidth and the transfer function e_2/e_v neglecting r_p. If e_1/e_v is on the order of unity, discuss why it is possible to use a triode for the grounded-cathode stage rather than a pentode.

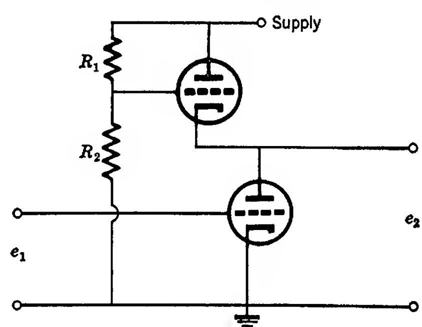


Fig. P.7.

7. Two triodes are connected as shown in Fig. P.7. Calculate the transfer function e_2/e_1.

8. If in the circuit of Fig. P.7 the transconductances of the two tubes increase slightly by the same amount with μ remaining constant (possibly due to a fluctuation in the filament supply voltage), what will be the effect on the transfer function?

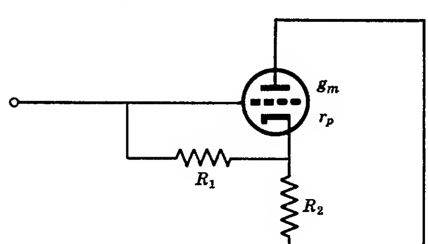


Fig. P.9.

9. Calculate the input impedance to the circuit of Fig. P.9.

10. A grounded-cathode pentode tube in a wide-band band-pass amplifier at 60 mcs has $g_m = 5000$ micromhos, a grid-to-cathode capacitance of 4 $\mu\mu$f, and a cathode lead inductance of 0.2 μh. What will be the input impedance if the plate load resistance at band center is (*a*) zero and (*b*) 2K?

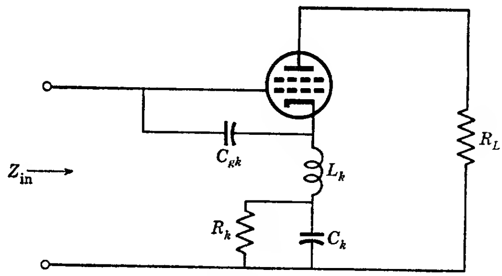


Fig. P.11.

11. Calculate the input impedance at the grid of the circuit of Fig. P.11 at the frequency where the impedance from cathode to ground is purely resistive. Also, determine this frequency.

12. Calculate the input impedances of the circuits of Fig. P.12 and discuss.

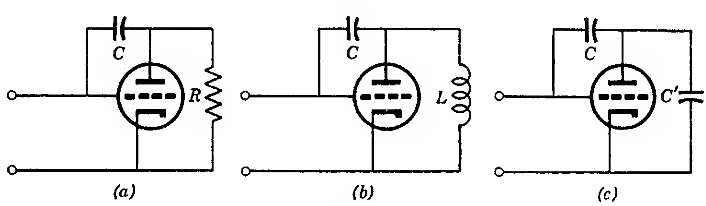


Fig. P.12.

13. Obtain the transfer functions e_2/e_1 for the circuits of Fig. P.13 and plot the locations of the p-z. Discuss.

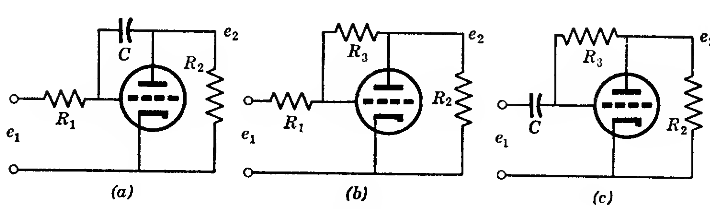


Fig. P.13.

14. Determine the gain at $\omega = 0$ and the bandwidth of the circuit of Fig. P.13*a* assuming $\mu = 50$, $g_m = 3000$ micromhos, $R_1 = 500$K, $R_2 = 25$K, and $C = 0.1$ μf.

15. Calculate e_a/e_{v1} for the circuit of Fig. 9.28 using $Z_k = R_k$, $Z_L = R_L$, and $e_{v2} = 0$. What will be the effect on the transfer function if the g_m of the two tubes changes in the same manner simultaneously?

16. At its operating point, a pentode draws 11 milliamperes (ma) plate current and 2 ma screen current at a screen voltage of 150. The tube is to have its cathode grounded except at very low frequencies. What cathode resistance is needed to give a grid bias of 2 volts? What screen dropping resistance is needed when the supply voltage is 250? What must be C_k (if C_{sg} is infinite) in order that the amplification be within 10 per cent of that at medium frequencies at 200 cps? What must be C_{sg} (if C_k is infinite) using this same criterion? Refer to Fig. P.16.

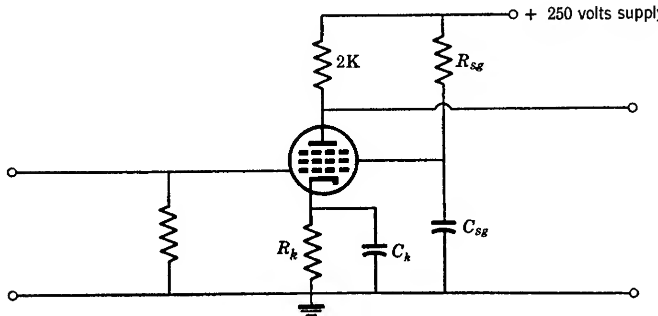


Fig. P.16.

17. A triode cathode follower has $g_m = 5000$ micromhos at $i_b = 12$ ma and $g_m = 2000$ micromhos at $i_b = 3$ ma with g_m approximately linearly dependent upon i_b for intermediate values. Plot the output impedance of the cathode follower as a function of i_b over this current range assuming μ is constant at 40. The value of the cathode resistor can be computed when the grid-leak resistor is returned to ground by assuming that a bias of 1 volt and 3 volts give plate currents of 12 and 3 ma respectively, with the bias for intermediate values varying linearly. Plot the value of the required cathode resistance as a function of the plate current. Refer to Fig. P.17.

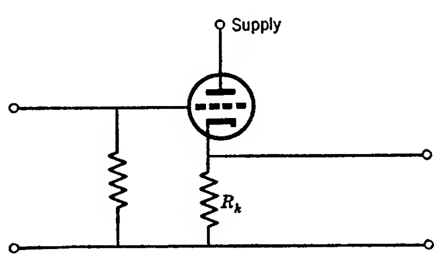


Fig. P.17.

18. Two grounded-grid stages are placed in cascade in a transmission line operating at high frequencies. The actual circuit (for a-c variations only) is shown in Fig. P.18a. It is desired to adjust G_1 so that the input impedance is equal to $Z_0 = R_0$ at the center frequency ω_0. Then, the equivalent circuit is that of Fig. P.18b. Determine e_L/e_0 if $g_m = 16{,}000$ micromhos, $\mu = 40$, $R_0 = 90$, $C = 15$ $\mu\mu$f, and the center

frequency is 100 mcs. On which tuned circuit is the transfer function most dependent?

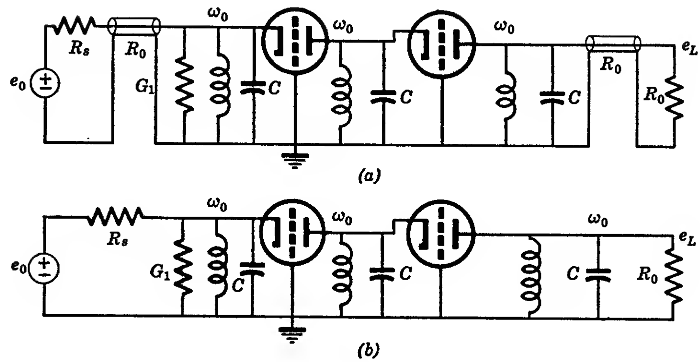


Fig. P.18.

19. Repeat the previous problem for the cascode amplifier of Fig. P.19 using the same parameters and compare with the results of Prob. 18.

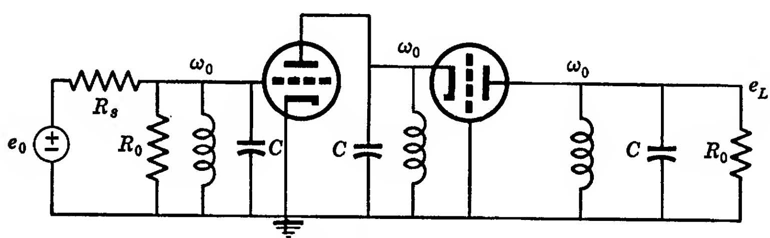


Fig. P.19.

10

Low-Pass Amplifiers

Armed with the concept of p-z and equipped with some knowledge of network design and of simple vacuum-tube circuits, we finally come to the point where the bits of information so far accumulated can be combined in order to design complete vacuum-tube systems. The road will not be at all difficult because it amounts to little more than putting together the pieces.

10.1 Networks and tubes

A vacuum tube is always operated in conjunction with a network. Sometimes two or more tubes may be part of the same basic network. Each basic network has a certain transfer function which is relatively simple and may represent one function in a cascaded arrangement of several functions.

Perhaps the one simple function is all that is required to perform some desired task. In that event, the equivalent circuit must be solved as a unit and its p-z or defining polynomials adjusted to agree with the desired collection of p-z or the coefficients of the desired polynomials (for example, a three-pole maximally flat function).

On the other hand, the over-all function may be quite complex. Several simple functions, each of whose characteristics are known, can be cascaded. The p-z of each simple function can be equated to a sub group of the total desired set of p-z and the type of circuit required as well as all element values can be obtained quickly.

The reader may ask why the over-all function need be complex when simpler functions of frequency might be adequate. The answer is basically a matter of gain. Most multitube linear devices perform the task of amplifying a signal covering a finite band of frequencies. But only a relatively small amount of gain with a sufficient bandwidth can be obtained with one tube so that several simple tube-network combinations in cascade may be needed in order to achieve the required gain along with an adequate bandwidth. However, the distributed capacitances always associated with tubes, as well as the coupling and by-pass net-

works needed so that correct operating voltages can be placed on the tubes, cause each tube-network combination to introduce p-z. The resulting cascaded system then shows a fairly large and irreducible number of p-z which must be positioned correctly in order to obtain a satisfactory over-all transfer function with a high over-all gain. The relatively large number of p-z may be employed to advantage in providing reasonably constant gain and linear phase over the band of frequencies of interest and a large rejection for frequencies outside the band. The larger the total number of p-z, the greater can be made this "off-channel" rejection; consequently, the greater can be made the discrimination of the system to signals occurring at frequencies outside the band of interest. If rejection without a large amount of gain is required, the network associated with each tube can be made more complex. This is, of course, preferable to building up a large number of p-z with an excessive number of tubes.

To achieve reasonable economy in power-supply requirements, cathode resistor and by-pass networks are often employed so that a small d-c potential difference can be placed between the grid and the cathode. Also, similar resistor and by-pass combinations may be placed at the screens of screen-grid tubes so that favorable screen voltages can be obtained without requiring separate power supplies for plates and screens. To keep the high d-c plate voltage of a tube from being applied to the grid or cathode of a subsequent tube, blocking capacitors must be introduced. All of these by-pass and blocking circuits are frequency sensitive at low frequencies. In fact, a blocking capacitor causes the gain to be zero at zero frequency because a zero is added to the transfer function at the origin of the p plane. The amplifier then actually displays a band-pass rather than a low-pass characteristic. However, if the low-frequency effects occur at frequencies that are very much smaller than the high-frequency cutoff point, the system can be classified as low pass and the high-frequency p-z can be ignored when studying low-frequency phenomena, and conversely. As long as the reduction in gain at low frequencies distorts the desired signal very little, no harm results from by-pass and blocking circuits.

If the transmission of direct current is important, blocking capacitors are not used, direct coupling (through resistors) being employed instead. Then there are no zeros at zero frequency. As would be expected, more complex power-supply systems are required for d-c amplifiers.

10.2 Building blocks and systems

A cascaded system composed of several tube-network units having individual transfer functions $f_1(p)$, $f_2(p)$, and so on is shown in Fig. 10.1.

Each of these tube-network boxes can be termed an "amplifier building block."

In obtaining the transfer function of a given block, the input impedance to the following block must be considered as part of the load impedance. For example, the distributed capacitance from grid to ground of a grounded-cathode amplifier should be considered to be part of the

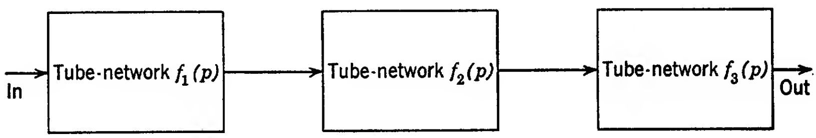


Fig. 10.1. The cascaded system.

load impedance of the preceding block and not associated with the block containing the tube that furnishes the capacitance.

If changes in the circuit elements of a block beyond the input to that block make no changes in the load impedance of the preceding block, the system consists of isolated tube-network stages. This kind of a system is the easiest with which to deal because each block can be designed as a separate entity with little regard for the effects of other blocks (other than a constant impedance attributed to them). An example of such a system is one consisting only of grounded-cathode tubes where the grid-to-plate capacitances are negligible and cathode lead inductance is unimportant.

When the system is not isolated in this sense, the changes referred to may be small enough to be ignored entirely or may vary so little that average values may be taken for input impedances. An example of this is a chain of grounded-cathode stages having small impedances between cathode and ground with the input impedance at the grid of each tube dependent upon grid-to-cathode capacitance.

At the extreme is the block whose input impedance is very sensitive to changes in its output circuitry as, for example, the grounded-grid amplifier or systems suffering from considerable Miller effect. For such systems, the separation into separate building blocks may not be justified and several building blocks may have to be lumped into one and the composite network solved with a single set of node or loop equations. However, this approach is difficult and can often be avoided through approximation. Usually, the input impedances to such blocks will strongly depend only upon immediately following blocks and the effects resulting from blocks two or three units down the chain can be ignored.

In the interests of simplicity, we shall concern ourselves only with isolated systems in this section. Procedures by which complicated trans-

fer functions can be built up with isolated building blocks can now be described in a general and rather complete manner. First, consider the system described by the p-z of Fig. 10.2a in which each pole and each zero is of order N. Clearly, a cascade of N identical blocks, each of

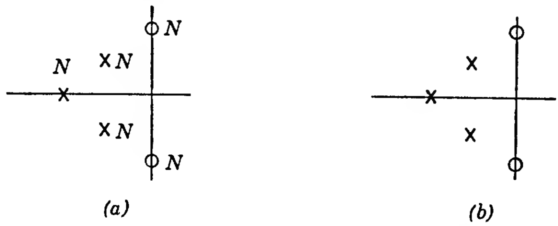


Fig. 10.2. A cascade of identical functions.

which has the p-z of Fig. 10.2b, achieves the desired result. Such systems are perhaps the most important for low-pass amplifiers.

Next, assume that it is desired to build up the maximally flat function of Fig. 10.3a. Of course, we could design a single network to realize the

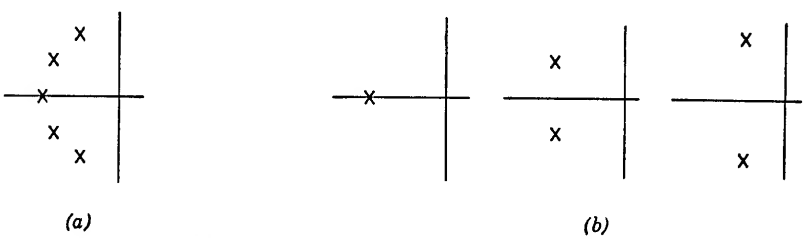


Fig. 10.3. Cascading nonidentical functions.

poles of Fig. 10.3a. However, we shall restrict our interest to cases where several stages are desired. To realize the function we could use three cascaded stages, each of which furnishes poles as shown in Fig. 10.3b. In other words, we require each block to furnish only a few of the p-z of the desired over-all function. Another way to realize the function of Fig. 10.3a is to use individual stages having p-z as shown in Fig. 10.4

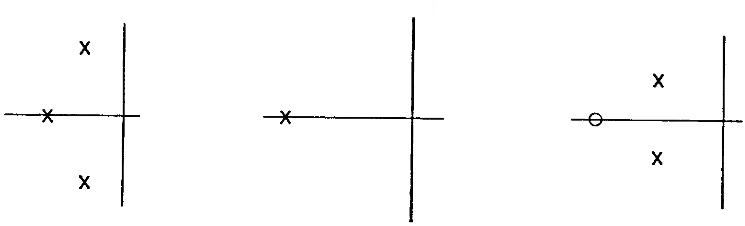


Fig. 10.4. An alternate cascade with p-z cancellation.

in which p-z cancellation is advantageously employed. The technique of p-z cancellation greatly increases the number of different kinds of building blocks of value for obtaining some given over-all transfer function.

Figure 10.5 shows an over-all transfer function in which a pole and a zero occur on the negative real axis in fairly close proximity. If the effect due to the zero is assumed to be cancelled by that of the pole,

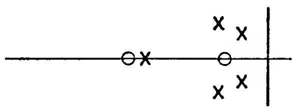


Fig. 10.5. A pole and a zero in close proximity.

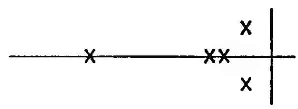


Fig. 10.6. A pole far out on the axis.

both can be neglected and the over-all transfer function thereby simplified. As a practical matter, p-z pairs in close proximity can usually be neglected. (The precision of a system is often not great enough to tell whether or not such a pair exists.)

Figure 10.6 shows a transfer function having one pole far out on the negative real axis. As far as its effect at frequencies near the rest of the p-z is concerned, it can be neglected and removed from the system. Unless the off-channel rejection at frequencies far from those near most of the p-z is important, its inclusion in the system is adding unnecessary complexity.

The precision with which the desired position of a pole or a zero must be realized is dependent upon the required precision of the over-all phase and gain variations with frequency. Happily, surprisingly large deviations can often be tolerated.

10.3 Isolated systems with two-terminal impedances

The most common amplifier building block consists of a grounded-cathode pentode or triode with but a single load impedance in the plate or cathode circuit as shown in Fig. 10.7 (which neglects coupling capacitors and power-supply voltages). The capacitances shown in Fig. 10.7 are the total capacitances at the corresponding points in the circuit. For example, C consists of the output capacitance of tube V_1, the input capacitance of tube V_2, and all extra capacitance brought about by circuit wiring and the tube sockets of V_1 and V_2. Similar capacitances exist between cathode and ground. The capacitance to ground at the output includes the capacitance of any load that the system drives. A typical minimum value for the interstage capacitance (neglecting Miller effects) is 15 $\mu\mu$f.

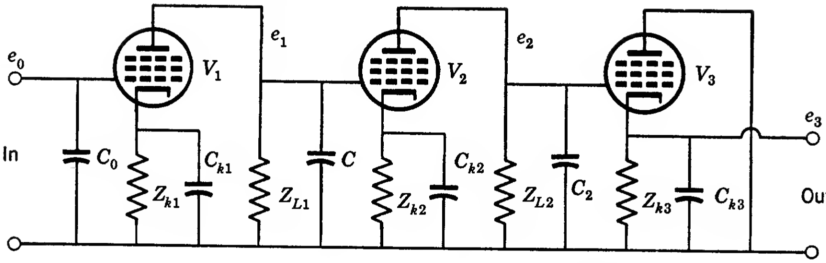


Fig. 10.7. An amplifier with two-terminal interstage networks.

The system of Fig. 10.7 is isolated if the grid-to-plate capacitance is negligible and the grid-to-cathode capacitance can be assumed to occur between grid and ground (as is always true when the cathode is by-passed to ground with a large capacitor).

We may compute the ratio e_1/e_0 of Fig. 10.7 very simply by removing the corresponding tube-network combination from the circuit and considering it as a separate entity. Similarly, we may compute e_2/e_1 and e_3/e_2. Multiplying these three functions together gives us the over-all transfer function.

Let us study the first tube of Fig. 10.7 along with its associated circuitry. The equivalent circuit is shown in Fig. 10.8*a*, in which the extra subscripts on Z_k and Z_L have been dropped for convenience. Let us

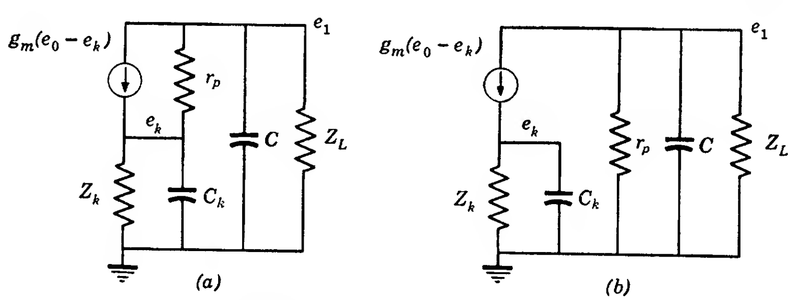


Fig. 10.8. Equivalent circuits.

approximate the circuit of Fig. 10.8*a* with that of Fig. 10.8*b*, in which the plate resistance r_p has been placed in parallel with the load rather than in parallel with the current generator. This approximation is virtually always justifiable with pentodes. The transfer function is found from Fig. 10.8*b* to be

$$\frac{e_1}{e_0} = \left(\frac{-g_m}{Y_L + pC + 1/r_p}\right)\left(\frac{Y_k + pC_k}{Y_k + pC_k + g_m}\right) \tag{10.1}$$

The first term in parenthesis of eq. 10.1 is related to the load impedance, whereas the second is related to the cathode impedance. The approximation of Fig. 10.8*b* has enabled us to factor the transfer function such that cathode and plate circuits can be considered separately. We may therefore write eq. 10.1 as

$$\frac{e_1}{e_0} = -g_m Z_{\text{in}} \left(\frac{Y_k'}{Y_k' + g_m} \right) \tag{10.2}$$

where Y_k' includes the capacitance C_k, and Z_{in} is the input impedance to the one-terminal-pair plate-load network of Fig. 10.9. If for the mo-

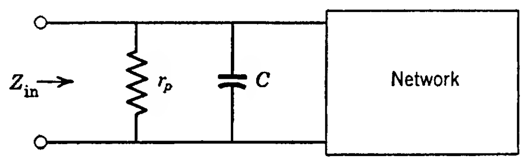


Fig. 10.9. The general two-terminal interstage network.

ment we neglect the factor containing Y_k' in eq. 10.2, we see that *the transfer function of an isolated stage having a two-terminal impedance is entirely dependent upon the input impedance to a network.* Therefore, the p-z of the transfer function are those of the network of Fig. 10.9.

Without further ado, we can set down a number of different simple networks usable as two-terminal load impedances and indicate typical p-z plots, as in Fig. 10.10. In all instances, the rules governing the p-z of input immittances (as well as transfer functions) must be obeyed, which means that the number of p-z can differ at most by unity, and so forth. In fact, if the capacitance C of Fig. 10.9 cannot be ignored, the transfer function $g_m Z_{\text{in}}$ will always have one more pole than zero.

Blocks that realize all-pole transfer functions are of considerable importance. In the type of circuit with which we are dealing here, obtaining one pole is possible with the *R-C* circuit of Fig. 10.10*a*. However, obtaining two poles without zeros by means of a single tube-network combination is impossible because of the laws governing the relative number of p-z of input immittances. This means, for example, that at high frequencies the magnitude of the transfer function will always decrease as $1/\omega$, never more. However, the *R-C* circuit of Fig. 10.10*a* in cascade with the shunt-peaked circuit of Fig. 10.10*b* can be made to furnish, for the *two* blocks, a pair of complex-conjugate (or real) poles if the pole of the *R-C* circuit is made to cancel the zero of the shunt-peaked circuit.

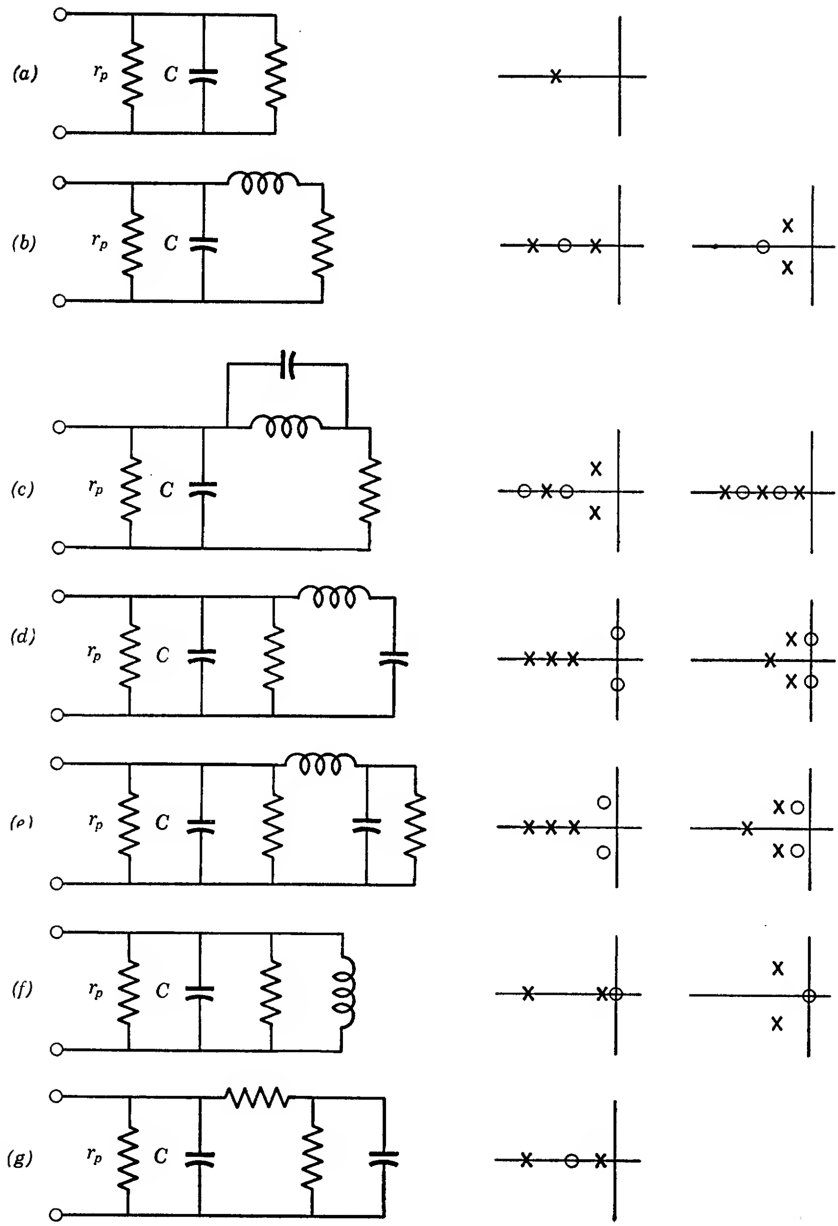


Fig. 10.10. Various two-terminal impedances.

Let us now consider the situation when Y_k' cannot be ignored, and in particular, when $Y_k' = G_k + pC_k$. Then, eq. 10.2 becomes

$$\frac{e_1}{e_0} = -g_m Z_{\text{in}} \left(\frac{p + G_k/C_k}{p + (G_k + g_m)/C_k} \right) \tag{10.3}$$

The factor in large parenthesis in eq. 10.3 has p-z as in Fig. 10.11*a*. If Z_{in} consists of R and C in parallel, the pole of Z_{in} will be as shown in Fig.

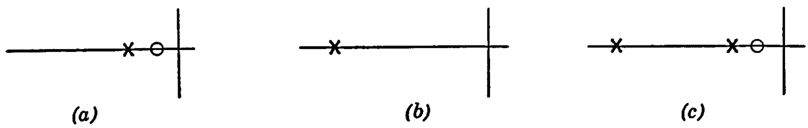


Fig. 10.11. Cathode and plate p-z.

10.11*b* and the complete transfer function e_1/e_0 will have p-z as in Fig. 10.11*c*. If the pole and zero introduced by the cathode circuit are close to the origin (as when C_k is large), it is primarily the low-frequency behavior of the amplifier stage that is affected by the cathode circuit. However, when C_k is small, the cathode circuit as well as the plate circuit will determine the high-frequency behavior. It is possible to adjust G_k and C_k so that the zero from the cathode circuit cancels the pole of the plate circuit as in Fig. 10.12. Then the bandwidth of the amplifier is increased as compared to that of the plate circuit alone. Cancellation in this manner is termed "cathode peaking." The advantage of cathode peaking is that low-frequency effects in the cathode due to a large bypass capacitor can be avoided with a small peaking capacitor. Also, some negative feedback can be realized.

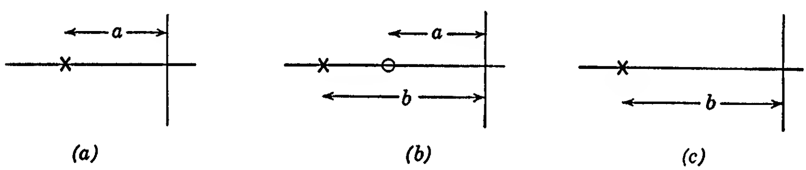


Fig. 10.12. Cathode peaking.

It should be observed that the gain at low frequencies with an R-C cathode circuit is

$$\left(\frac{e_1}{e_0}\right)_{\omega=0} = -g_m R_{\text{in}} \frac{G_k}{G_k + g_m} \tag{10.4}$$

where R_{in} is the value of Z_{in} at $\omega = 0$. Thus, the gain is less than it

would be if the cathode were grounded so that $R_k = 0$. With a large C_k, the cathode is essentially grounded at all but very low frequencies. The "cathode degeneration" that occurs when employing cathode peaking will not, however, reduce the net stage gain for a fixed bandwidth because the peaking permits the use of a larger plate load resistor than would otherwise be possible.

The screen circuit of a pentode can be treated much like the cathode circuit because g_m is replaced with g_m times a factor containing only screen circuit parameters, $g_m/(1 + Z_s/r_s)$, which expresses exactly the same type of behavior as that resulting from a finite cathode impedance; Z_s is the analog of Z_k and $1/r_s$ is the analog of g_m.

Finally, let us study the properties of the cathode follower. Tube V_3 in Fig. 10.7 is a cathode follower, the transfer function of which is

$$\frac{e_3}{e_2} = \frac{\mu}{1+\mu}\left(\frac{Z_k}{\dfrac{r_p}{1+\mu} + Z_k}\right) \cong \frac{\mu}{r_pY_k + \mu} = \frac{g_m}{g_m + Y_k} \tag{10.5}$$

where Z_k is the impedance of the parallel combination of Y_{k3} and C_{k3} in Fig. 10.7, and where the approximation is valid for large μ. If, as is most frequently the case, $Y_k = G_k + pC_k$ (R and C in parallel), we get

$$\frac{e_3}{e_2} = \frac{g_m/C_k}{p + [(g_m + G_k)/C_k]} = \frac{g_m/C_k}{p + B} \tag{10.6}$$

where the approximation of eq. 10.5 is assumed and where B is the half-power bandwidth in radians. Due to its low output impedance, the bandwidth of the cathode follower is quite large. The voltage gain at $\omega = 0$ is $g_m/(g_m + G_k)$ which, for $g_m \gg G_k$, is nearly unity.

The transfer function of the grounded-cathode plate-loaded circuit with $Z_k = 0$ and with $Y_{\text{in}} = G + pC$ was found to be

$$\frac{e_1}{e_0} = -g_mZ_{\text{in}} = \frac{-g_m/C}{p + G/C} = \frac{-g_m/C}{p + B} \tag{10.7}$$

which is normally adjusted to have a larger gain but a smaller bandwidth than the cathode follower. If G is increased until the voltage gain at $\omega = 0$, g_m/G, is about unity, the bandwidth will be about the same as that of the cathode follower (assuming the shunt capacitances are about the same). Of course, negative feedback is not provided as it is with the cathode follower. It would appear that a statement of bandwidth or gain is meaningless if considered by itself. However, the product of

bandwidth and gain, the "gain-bandwidth product," is quite meaningful. We shall study this in detail later.

10.4 Isolated systems with four-terminal impedances

A circuit employing a four-terminal network as an interstage network is shown in Fig. 10.13. At the input to the network, the output capacitance and plate resistance of V_1 appear. However, the input capacitance of V_2 does not appear at that point but rather at the output of the network. The four-terminal network achieves a splitting of the total

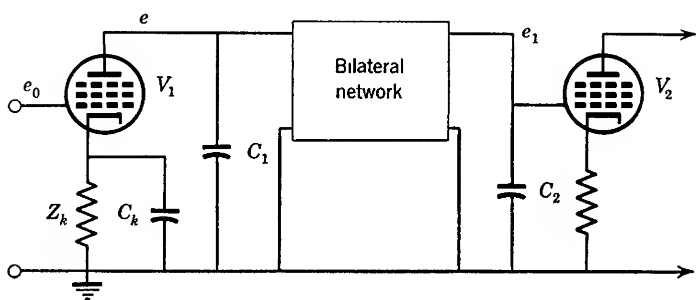


Fig. 10.13. Amplifier using a four-terminal interstage network.

interstage capacitance. We might expect that stage gain-bandwidth products of perhaps twice those obtainable with two-terminal impedances can be achieved.

The only difference between the transfer function e_1/e_0 for Fig. 10.13 and that for two-terminal impedances is that the transfer impedance Z_T of the network is substituted for the input (or load) impedance. The transfer function analogous to the approximation of eq. 10.2 is

$$\frac{e_1}{e_0} = -g_m Z_T \left(\frac{Y_k'}{Y_k' + g_m}\right) \tag{10.8}$$

With the two-terminal load impedance, we had $e_1/e_0 = -g_m Z_{\text{in}}$ times the same factor containing Y_k' (where it is assumed that Y_k' in eq. 10.8 includes shunt capacitance).

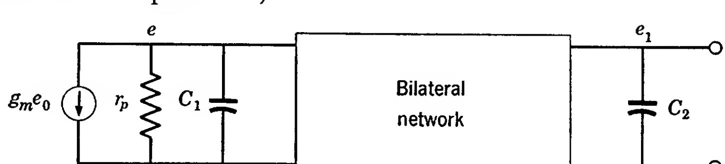


Fig. 10.14. Equivalent circuit for the four-terminal interstage.

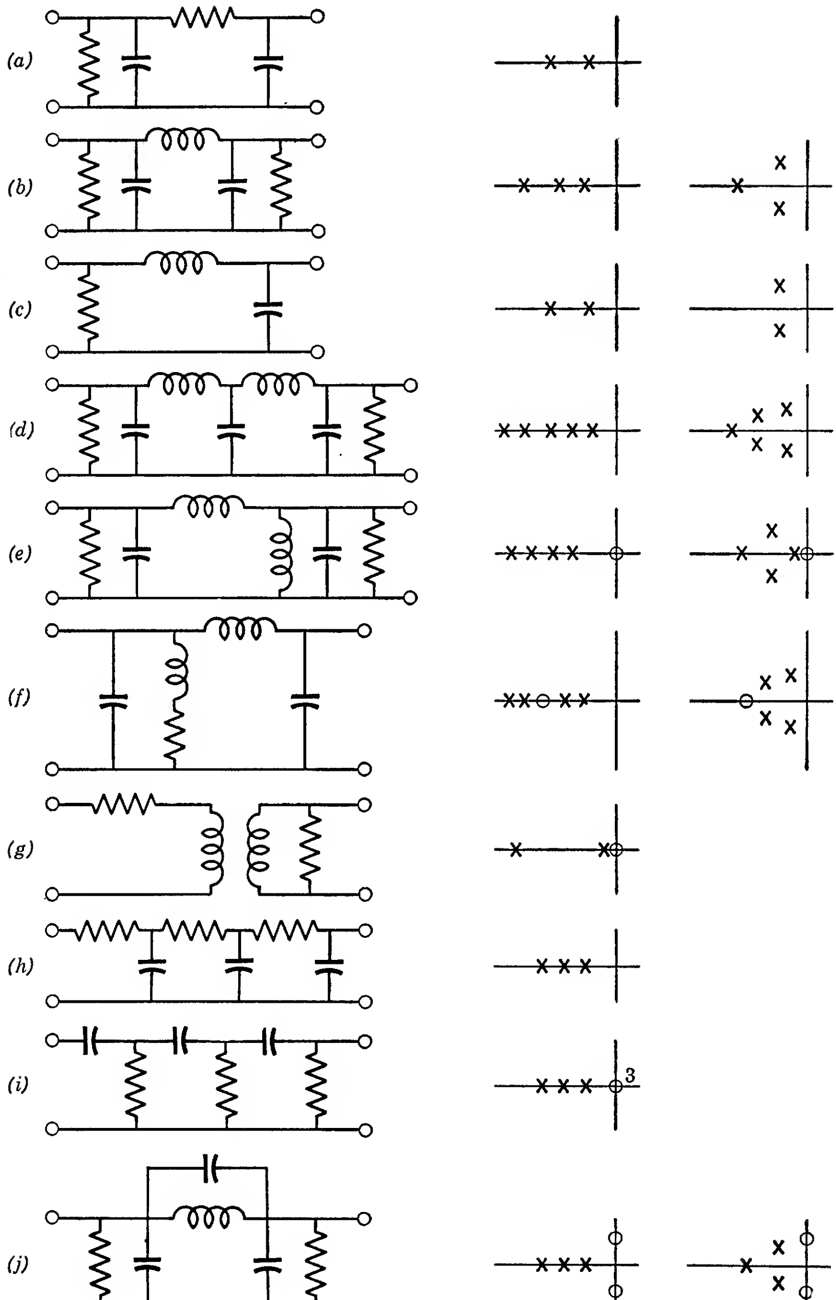


Fig. 10.15. Various four-terminal networks.

It was pointed out in Chap. 3 that transfer impedances are much less restricted than input impedances. In particular, functions with two or more poles than zeros can be realized with a single four-terminal interstage network.

The remaining discussion here need not be lengthy. All we must do is study particular transfer impedances as we studied input impedances before. The basic four-terminal network containing irreducible elements is shown in Fig. 10.14 from which can be calculated the transfer impedance Z_T.

A few specific networks are shown in Fig. 10.15 and typical p-z locations for the transfer impedances indicated. In some circuits of Fig. 10.15, stray capacitances have been ignored. The circuit of Fig. 10.15*f*, when driven with a current generator, is the most efficient wide-band interstage network employed in practice, the "series-shunt-peaked" network.

10.5 Transformer coupling

A transformer with a large coefficient of coupling is frequently employed as an interstage or output network in audio amplifiers. It is also used in wide-band systems designed for the amplification of fast signals such as pulses. Basically, the transformer is a four-terminal interstage network having a low-frequency behavior that to a certain extent is separable (analytically) from the high-frequency behavior. There is no fundamental difference between audio and pulse transformers and ordinary 60-cycle power transformers. The only significant differences are the volume of the iron core material and the thickness of the core laminations. Core volume may be somewhat smaller in audio transformers and considerably smaller in pulse transformers, depending on the minimum acceptable magnetizing inductance. Lamination thicknesses are made as small as is practical, particularly in pulse transformers, in order to minimize eddy-current losses.

Most low-pass networks do not permit a conjugate match between different source and load impedances to be realized and consequently are disadvantageous in many instances where high efficiency is desired. The only way high efficiencies can be obtained in low-pass systems is to use near-ideal transformers.

Let us study the transformer operating between source and load resistances R_s and R_L through a development of its equivalent circuit. The reader can follow the step-by-step reduction of Fig. 10.16 in order to study the manner of pulling out and accounting for the various parameters of a practical transformer: the primary and secondary stray capacitances C_1 and C_2, leakage reactances L_1 and L_2, wire resistances R_1 and

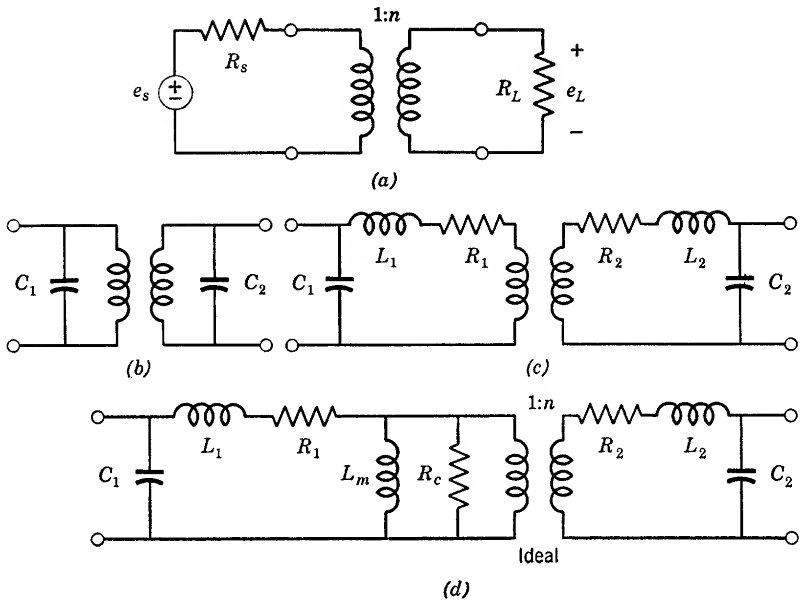


Fig. 10.16. Transformer reduction.

R_2, and the magnetizing inductance L_m and core-loss representation R_c, both referred to the primary. We are left with a network of simple circuit elements and an *ideal* transformer.

Now we proceed to refer all quantities in the secondary (load side) of the ideal transformer to the primary in a step-by-step fashion. This can be done for the load resistor R_L as well so that we need study only

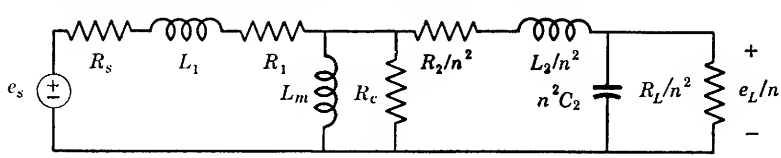


Fig. 10.17. Final transformer equivalent.

the referred values of load voltage and current rather than their actual values. The completely reduced circuit is shown in Fig. 10.17.

The low-frequency behavior is determined largely by the short-circuiting effect of the magnetizing inductance L_m; therefore L_m is normally made relatively large. The high-frequency behavior is determined largely by the leakage reactances, which must thus be made relatively small. If the ratio of frequencies at which leakage and magnetizing inductances are important is to be large (hundreds in audio trans-

formers), then the ratio of magnetizing inductance to leakage inductance must be large, which in turn requires that the coefficient of coupling be fairly close to unity.

At low frequencies, leakage reactances and stray capacitances are relatively unimportant compared to the magnetizing inductance. Therefore, the circuit valid at low frequencies can be drawn as in Fig. 10.18,

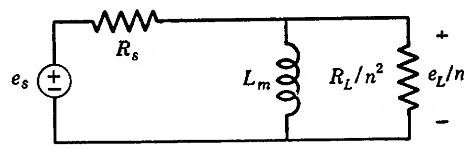


Fig. 10.18. Transformer equivalent at low frequencies.

where R_1, R_2, and R_c have been ignored for simplicity. The transfer function of this circuit has a zero at the origin and a single pole not too far out on the negative real axis and is the same as that of a simple R-C coupling circuit.

At high frequencies, the reactance of L_m is so large that it can be neglected. Therefore, the equivalent circuit becomes that shown in

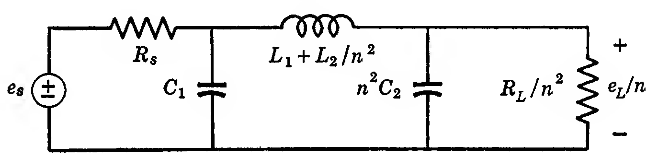


Fig. 10.19. Transformer equivalent at high frequencies.

Fig. 10.19, where R_1, R_2, and R_c have again been neglected for simplicity. The reader should be quite familiar with the circuit of Fig. 10.19 by now.

The p-z plot of the transfer function of the transformer-coupled circuit will have a form typically like one of those shown in Fig. 10.20. A trans-

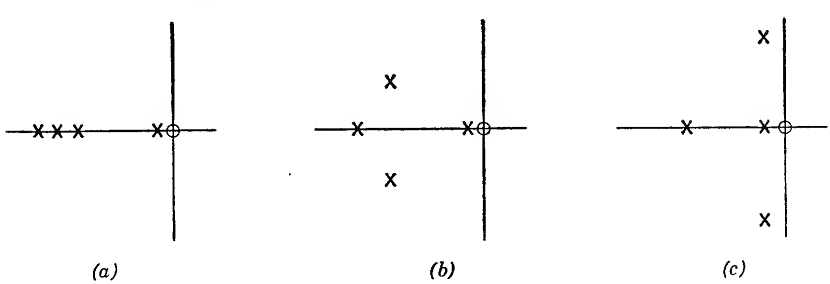


Fig. 10.20. Transformer p-z plots.

former with a large magnetizing inductance is entirely equivalent to a series-peaked type of low-pass circuit in cascade with an R-C coupling circuit and an ideal transformer. These general characteristics are about the same even when R_1, R_2, and R_c are not neglected.

If one of the two shunt capacitances C_1 or C_2 is neglected (as is sometimes a valid approximation when the step-up or step-down turns ratio is large), the number of high-frequency poles is reduced from three to two. If both capacitances are neglected, there is only one high-frequency pole and the transfer function cannot be made to show a high-frequency peak.

Frequently, the plate current of a tube driving a transformer flows through the primary winding. This current tends to saturate the core with d-c flux and, if too large, greatly reduces the magnetizing inductance and increases the leakage inductances. It is always important not to exceed the manufacturer's ratings in this regard. When a transformer is in a push-pull circuit with the power-supply connection brought to a center tap on the primary, the direct currents to the two tubes result in opposite magnetizing effects so that there is no d-c magnetization of the core.

10.6 Building blocks with feedback

Certain feedback arrangements are capable of realizing desirable transfer functions. Of special interest is the "feedback pair" shown in Fig. 10.21, which furnishes two complex-conjugate poles and no zeros.

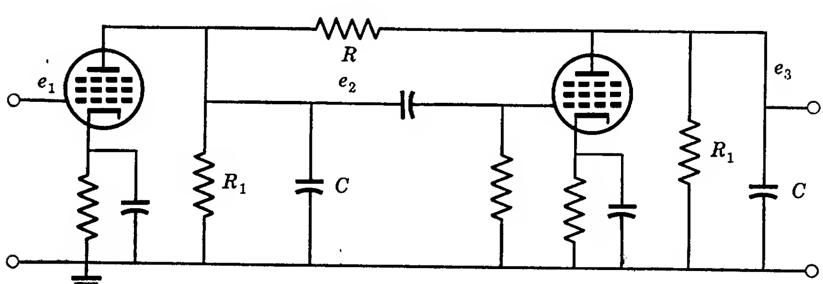


Fig. 10.21. The feedback pair.

Feedback triples and so on can also be built in order to realize various all-pole functions where the poles are complex. It is an easy matter to change bandwidth by modifying the feedback resistor R.

It should be observed that the feedback pair is just one of a multitude of building blocks that requires the use of more than one tube. Since it is a simple one, it is best studied in this chapter, although more com-

plex feedback circuits will be studied in later chapters. For some reason, the feedback pair has not been nearly as popular as is warranted by its excellent characteristics and utter simplicity. (It gives poles having imaginary parts without the use of inductance.)

Assume that the two tubes of Fig. 10.21 are identical and that r_p can be neglected ($r_p \gg R_1$). Also, assume that the plate load impedances are identical. We shall confine our interest to the high-frequency behavior and therefore assume that the cathodes are essentially at ground potential (large C_k). With these assumptions, the equivalent circuit is that of Fig. 10.22.

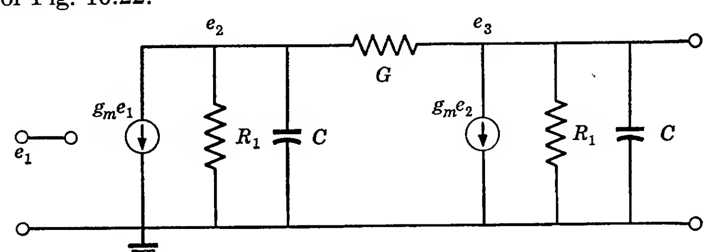


Fig. 10.22. Equivalent circuit of the feedback pair.

The equations at the nodes where voltages e_2 and e_3 exist are

$$\begin{aligned} (G + G_1 + pC)e_2 - Ge_3 &= -g_m e_1 \\ -Ge_2 + (G + G_1 + pC)e_3 &= -g_m e_2 \end{aligned} \tag{10.9}$$

From these equations, we obtain the transfer function as

$$\frac{e_3}{e_1} = \frac{[g_m(g_m - G)]/C^2}{p^2 + p[2(G + G_1)/C] + \{[G(g_m - G) + (G + G_1)^2]/C^2\}} \tag{10.10}$$

which has a pair of complex-conjugate poles at

$$p_1, p_1^* = -\frac{G + G_1}{C}\left(1 \pm j\frac{\sqrt{G(g_m - G)}}{G + G_1}\right) \tag{10.11}$$

Often, $g_m \gg G$ permitting G to be neglected in the factor $(g_m - G)$. If also $G \gg G_1$, we have

$$p_1, p_1^* = -\frac{G_1}{C}\left(1 \pm j\frac{\sqrt{g_m G}}{G_1}\right) \tag{10.12}$$

It is evident that the feedback resistor G can be varied to change the imaginary part of the pole position without affecting the real part. This causes the bandwidth to vary.

10.7 Low-frequency behavior

A typical circuit for a tube with screen and cathode by-passing is shown in Fig. 10.23, where the supply bus can be assumed to be at ground potential as far as alternating voltages and currents are concerned. To obtain the low-frequency circuit, all distributed and other

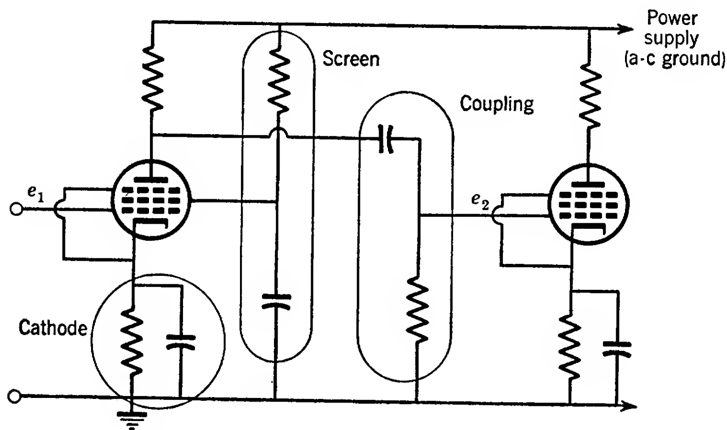


Fig. 10.23. By-passing and coupling circuits.

small capacitances are ignored and all small inductances are considered to be short-circuited or are replaced with loss resistors.

The cathode, screen, and coupling networks supply p-z to the transfer function e_2/e_1 as shown in Figs. 10.24*a*, *b*, and *c* respectively. All of these are characteristic of high-pass circuits; thus, each plays a part in

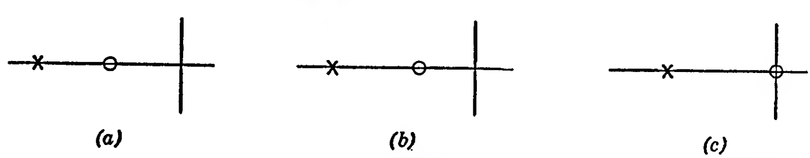


Fig. 10.24. Low-frequency p-z. (*a*) Cathode. (*b*) Screen. (*c*) Coupling.

reducing the gain at low frequencies. By making the product of the resistance and capacitance (the "*R-C* product" or "time constant") of each circuit larger, the pole, in each case, is moved closer to the zero and the gain at low frequencies is increased accordingly.

Sometimes a "compensating circuit" helps bring up the low-frequency response. Such a circuit in the plate load appears in Fig. 10.25; at high frequencies, the impedance of C is so small that point "x" can be assumed to be at ground potential. It should be observed that this cir-

cuit can entirely remove the effects of the screen or the cathode by-pass network by p-z cancellation. It is widely used to improve the long-time response of pulse amplifiers. The optimum design criterion for low frequencies (to be derived in Sec. 10.11) is as follows: when the sum of the low-frequency pole positions is equal to the sum of the low-frequency

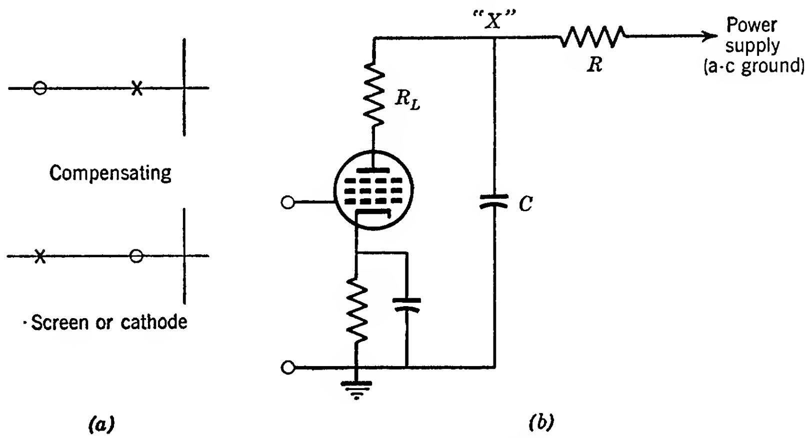


Fig. 10.25. Low-frequency compensating circuit.

zero positions, optimum low-frequency compensation (in one sense) has been achieved.

10.8 The gain-bandwidth product

The grounded-cathode tube with an R-C plate load has a voltage gain at $\omega = 0$ of g_m/G and a bandwidth of G/C, where G and C are the input conductance and capacitance of the plate load impedance. Thus, the gain-times-bandwidth product is

$$(GB)_{\text{stage}} = (g_m/C)(G/C) = g_m/C \tag{10.13}$$

This factor is dependent only upon the tube transconductance and (irreducible) stray capacitance C and appears to be a measure of tube performance. The minimum possible value of C is the sum of the output capacitance of one tube plus the input capacitance of the following tube or the load that the tube drives. The factor varies over quite large ranges, being about 400×10^6 radians per second for typical tubes designed for wide-band amplifiers. The minimum possible value of the capacitance is entirely dependent upon the tube and the tube alone. When the tube is in a circuit, the minimum possible value will be some-

what larger by virtue of the stray and other capacitance introduced by the wiring (or purposely introduced).

If N building blocks are cascaded in an isolated system and the over-all gain is G_T at $\omega = 0$ (considering only the high-frequency p-z), a geometric mean stage gain can be defined as $(G_T)^{1/N}$. The GB product applied to the complete system is defined as the *mean* gain times the *over-all* bandwidth B_T as

$$(GB)_{\text{system}} = (G_T)^{1/N} B_T \tag{10.14}$$

If all N stages consist of simple and identical R-C networks, the GB product per stage is g_m/C and the mean stage gain is g_m/G. The over-all bandwidth is less than that of each stage because of the bandwidth-narrowing equation. Then

$$(GB)_{\text{system}} = (g_m/C)(2^{1/N} - 1)^{1/2} \tag{10.15}$$

When many stages are employed, the system gain-bandwidth product becomes much less than that for a single stage. This means that each stage must have a much wider bandwidth than that desired for the entire system; hence, tubes must be liberally added to bring up the gain to that required because increasing the bandwidth of a stage also reduces its gain. It should be apparent that it is impossible to build an amplifier with a gain greater than unity using R-C plate loads if the required bandwidth is greater than the g_m/C of the tubes to be employed.

As an example, assume that an amplifier having a gain of 10^4 and bandwidth of 60×10^6 radians per second is to be designed with tubes having a g_m/C of 400×10^6 radians per second and with identical R-C plate loads. If ten stages are used, the narrowing factor is 0.268. Thus, each stage must have a bandwidth of $60/0.268 = 224 \times 10^6$. This means that the stage gain is $400/224 = 1.78$. Ten such stages give a gain of 320, which is not nearly enough. So, we try 14 stages in an attempt to solve the problem. The narrowing factor becomes 0.225. This leads to a mean stage gain of 1.5; thus, the over-all gain is 290 which is even less. Fewer stages actually give more gain! It appears that the design problem cannot be solved with R-C interstage networks. The situation would be even worse had we used nonidentical R-C networks. In fact, when the mean stage gain is $(e)^{1/2} = 1.65$, an amplifier having identical R-C plate loads will have the maximum possible over-all gain for a given over-all bandwidth.

We have demonstrated that, with a given number of tubes and with a given bandwidth, a certain gain may be achieved. With the same number of tubes and the same over-all bandwidth, the question then arises as to the possibility of obtaining more over-all gain.

The GB product of the R-C network is g_m/C. Some other two-terminal networks, when used with the same tube, have GB products larger than g_m/C. This helps to increase the over-all gain. In addition, some of the circuits are not as severely affected by bandwidth narrowing as the R-C circuit, which gives an even greater improvement. For example, a cascade of identical shunt-peaked circuits adjusted to be maximally flat can solve the design problem posed in the example. A cascade of series-shunt-peaked stages (four-terminal network) results in a very efficient solution.

A general expression valid whenever identical single-tube stages are cascaded is

$$(GB)_{\text{system}} = (g_m/C)(k)(F_N) \tag{10.16}$$

where g_m/C is the figure of merit for the tube as has previously been described, k is the advantage factor of the interstage circuit relative to the R-C two-terminal network, and F_N is the narrowing factor for N stages. It must be assumed that the individual stages are relatively flat with the gain dropping off smoothly with frequency if eq. 10.16 is to be valid; if a peak in the amplitude response exists, cascading will magnify the peak and eq. 10.16 will not be accurate.

The factor k for R-C stages is unity. For other two-terminal impedances it may be somewhat larger, up to about two. For four-terminal interstage networks, the relative advantage factor will often be about two and may be as high as about four. The bandwidth-narrowing factor depends on the form of the interstage network. For some circuits it may decrease rapidly and for others relatively slowly as the number of cascaded stages N is increased. Clearly, F_N cannot exceed unity.

One of the more efficient four-terminal interstage circuits applicable to a cascaded system of identical stages is the series-peaked circuit adjusted to be maximally flat. It has three poles. The factor k for this circuit (assuming that the output capacitance of the tube rather than the input capacitance of the following tube is the basically irreducible element) can be as large as 2.0, depending on the ratio of tube output to input capacitances. The big advantage comes because the narrowing factor is small. With this circuit in the previous example (and for $k = 2$), the maximum stage gain-bandwidth product (g_m/C times the advantage factor k) is 800×10^6. For five stages, the narrowing factor is $(2^{1/5} - 1)^{1/6} = 0.727$. Thus, each stage must have a bandwidth of $60/0.727 = 82.5 \times 10^6$. The mean stage gain is therefore $800/8.25 = 9.7$ giving a total gain of $(9.7)^5 = 87{,}500$, which is more than that required by the example. Four stages would have been adequate. The use of this circuit to obtain a design for the amplifier of the example

furnishes a rather striking comparison to the abortive attempt with simple R-C networks.

Another good solution to the amplifier problem of the example is had with three feedback pairs (six tubes) which has the advantage of requiring no inductance, merely simple (and cheap) resistors. If each pair is maximally flat and identical, the mean stage gain-bandwidth product (applicable to each pair) is $g_m/C = 400 \times 10^6$. The narrowing factor is $(2^{1/3} - 1)^{1/4} = 0.715$. The bandwidth required per pair is $60/0.715 = 84 \times 10^6$, which causes the mean stage gain to be $400/84 = 4.76$. Thus the over-all gain is $(4.76)^6 = 12{,}000$, which is also a solution to the problem. Extra tubes are required as compared to the amplifier employing series-peaked circuits but several inductances are made unnecessary, the adjustment of the amplifier is simpler, and some negative feedback is provided.

In all previous examples, the problem of obtaining a gain of 10^4 rapidly becomes more difficult as the required bandwidth exceeds 60×10^6 radians per second by any appreciable amount. When the required bandwidth is on the order of 120×10^6, it appears economically impractical to build high-gain systems consisting of identical cascaded stages. It can be concluded that a system of cascaded identical stages is limited when high gain must be obtained with large bandwidths, even though more advantageous circuits are employed.

Another and quite dissimilar method exists for building wide-band amplifiers; that is, all the stages can be made different so that the entire amplifier has a single n-pole maximally flat or equal-ripple or similar response. For example, a six-pole maximally flat function can be made from three feedback pairs, each pair furnishing one pair of complex-conjugate poles in the six-pole function. Alternately, three shunt-peaked stages followed by three R-C stages to cancel the zeros can be made to provide a six-pole maximally flat function. Actually, a considerable variety of circuits can be operated in tandem to give the desired function, each stage contributing a few of the total number of p-z.

Since each stage is quite different, it is not too easy to define a stage gain-bandwidth product. For example, the stage that furnishes the end poles of a maximally flat all-pole function, when considered by itself, may be more characteristic of a band-pass than a low-pass function. Similarly, a stage that contributes but a single pole will not look maximally flat at all. However, we can still define a mean stage gain and an over-all bandwidth as before. The main effect on the system gain-bandwidth product is to remove the narrowing factor. For a relative advantage factor k of unity, we get

$$(GB)_{\text{system}} = g_m/C \tag{10.17}$$

which applies when the amplifier consists of feedback pairs or shunt-peaked and R-C combinations to give an n-pole (no zero) maximally flat transfer function.

If an amplifier having a bandwidth any smaller than g_m/C is to be built, any desired gain can be achieved by making the entire amplifier have a single maximally flat transfer function. For our example, we can use two feedback pairs and an R-C stage to give a five-pole maximally flat function with five tubes. The mean stage gain is $400/60 = 6.67$, giving a total gain of $(6.67)^5 = 13{,}000$. If the pairs were arranged to give a medium-ripple Chebyshev response, the gain could be made considerably greater. When a maximally flat system is converted to one with a Chebyshev behavior, all the poles are moved slightly closer to the origin without changing the bandwidth. This increases the gain of each stage and results in

$$(GB)_{\text{system}} > g_m/C \tag{10.18}$$

with the inequality growing as the size of the ripples in the pass band is increased.

Some interstage networks, when used to provide part of an over-all maximally flat or equal-ripple function, yield an advantage factor k greater than unity. For example, a series-peaked circuit followed by a shunt-peaked circuit can be made to give a four-pole maximally flat function with an appreciable advantage factor.

The types of building blocks required in an amplifier are largely dependent on the required over-all gain-bandwidth product. If this product is large, maximally flat or equal-ripple functions may be best. If small, a cascade of identical stages will probably be most efficient. The number of tubes required can be determined trying various systems on paper, the solution requiring the fewest tubes being the most efficient. However, if the number of tubes required when a simple interstage system is used is not much greater than the number required for a more sophisticated system, the simplicity of the former may actually cause it to be a better solution to the problem.

In this connection, a "valid law" relating to psychology can be stated. That is, any system subject to tuning or trimming adjustments by relatively untrained individuals should be as simple as possible because the individuals will be likely at some time to try to make adjustments. These persons should not be given the opportunity to change anything but tubes; hence, a multitude of tubes is not too serious from this point of view, but complex interstage networks and critically tuned circuits are serious. One approach to the problem of the ambitious neophyte is to seal the entire amplifier so tight in a container that even the most persistent individual will have trouble dissecting it.

10.9 Some specific design comparisons

Let us compare interstage schemes for several amplifiers where each amplifier has four tubes (as an example). The comparison will be made on the basis of the over-all gain for a fixed over-all maximally flat bandwidth B. In the process the element values will be found, which are the important aspects of the high-frequency design. It will be assumed that the tubes are identical; therefore, the transconductance g_m and the total interstage capacitance C of each of the four stages will be assumed to be the same.

The very great emphasis that seemingly has been placed on maximally flat functions should not be taken too literally. These functions are merely representative. More important, they have simple mathematics, which makes them ideal for comparison purposes. If, say, a linear-phase function were actually desired, a comparison yielding relative efficiency based upon the maximally flat function would be nearly correct for the linear-phase function, although specific numerical values of bandwidth and gain might differ.

First, consider four identical R-C stages. The mean gain is g_mR and the over-all bandwidth is $(2^{1/4} - 1)^{1/2}/RC = 0.435/RC$. Thus, the mean gain times the over-all bandwidth is $0.435g_m/C$. We shall compare all other designs to this four-stage R-C amplifier. For a given C and B, the R-C stage has only one design parameter R, which must be $R = 0.435/BC$.

Let us now try a cascade of four identical shunt-peaked circuits, each of which is maximally flat. From Chap. 5, the general narrowing factor for N stages is

$$\left(\frac{B_N}{B}\right)^2 = \frac{2^{1/N} - 1 + \sqrt{(2^{1/N} - 1)^2 + 4\delta^2(2^{1/N} - 1)}}{1 + \sqrt{1 + 4\delta^4}} \tag{10.19}$$

where δ is the zero distance when the poles are normalized to lie on the unit circle. If $\delta = 2\alpha = (1 + 2^{1/2})^{1/2} = 1.554$ is used in eq. 10.19 for $N = 4$, we get $B_N/B = 0.623$.

The maximally flat shunt-peaked network ($\delta = 2\alpha$) normalized to $R_0 = 1$ and with poles on the unit circle has a bandwidth of 1.11. Also $L_0 = 0.643$ h and $C_0 = 1.554$ f (as in Chap. 5). In the circuit we must design here, the required over-all bandwidth is B. The bandwidth per shunt-peaked stage must therefore be $B/0.623 = 1.605B$. Thus we must unnormalize the bandwidth of the circuit by the factor $1.605B/1.11 = 1.44B$ for bandwidth. If C is the actual interstage capacitance, then $C = 1.554/1.44BR$, or $R = 1.08/BC$. Then $L = 0.643R/1.44B = 0.446R/B$.

We found the resistance for four R-C stages to be $0.435/BC$. Thus, the ratio of the resistance of each of the four shunt-peaked stages to that of each of the four R-C stages is $1.08/0.435 = 2.48$. Therefore, the four shunt-peaked stages will have a gain $(2.48)^4 = 38$ times as large as the four R-C stages for the same over-all bandwidth.

Now let us try four identical maximally flat series-peaked stages. For unity bandwidth, $R = 1$, $C_1 = \frac{1}{2}$, $L = \frac{4}{3}$, and $C_2 = \frac{3}{2}$. The narrowing factor is $(2^{1/4} - 1)^{1/6} = 0.757$. For an over-all bandwidth B, the

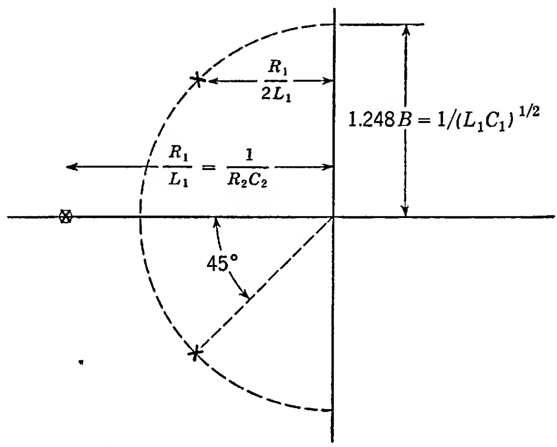


Fig. 10.26. Shunt-peaked and R-C p-z.

stage bandwidth must therefore be $B/0.757 = 1.32B$. In order to make this problem interesting, assume that the total unavoidable interstage capacitance C is contributed in a 1:2 ratio by stray and tube capacitances. Thus, the output capacitance of one tube is $C/3$ and the input capacitance to the next tube is $2C/3$. However, the maximally flat series-peaked circuit requires a 1:3 capacitance ratio, which means that the larger capacitance must be increased by padding from $2C/3$ to C, which is the original total unavoidable interstage capacitance. We therefore find $C = 1.5/1.32RB$, or $R = 1.135/BC$. Also, $L = 1.33R/1.32B$. A single series-peaked stage therefore has a gain $1.135/0.435 = 2.61$ times that of a single R-C stage for the same over-all bandwidth to give an over-all gain 46.5 times as large. Had the tube capacitance ratio been 1:3 without padding, the relative gain would have been $2.61 \times 1.5 = 3.91$ for an over-all gain advantage of 230.

Let us now try some composite systems. First, we shall use a shunt-peaked stage in cascade with an R-C stage to give a two-pole maximally flat function. Two such pairs of stages give a cascade of two two-pole

functions with a narrowing factor of $(2^{1/2} - 1)^{1/4} = 0.802$. Each pair must therefore have a bandwidth of $B/0.802 = 1.248B$, where B is the over-all bandwidth. The relevant p-z plot for each of the two pairs is shown in Fig. 10.26, where subscript 1 applies to the shunt-peaked stage and subscript 2 to the R-C stage. Let $C_1 = C_2 = C$ be the total interstage capacitance. Then $(L_1C)^{1/2} = 1/1.248B$ and $L_1 = 0.643/B^2C$. From the pole dimensions, $R_1 = 1.135/BC$ and $R_2 = 0.566/BC$. The over-all gain is proportional to $(1.135)^2(0.566)^2 = 0.414$. The over-all gain of four R-C stages with the same over-all bandwidth is proportional to $(0.435)^4 = 0.036$. Thus, this system has 11.5 times as much gain as four R-C stages.

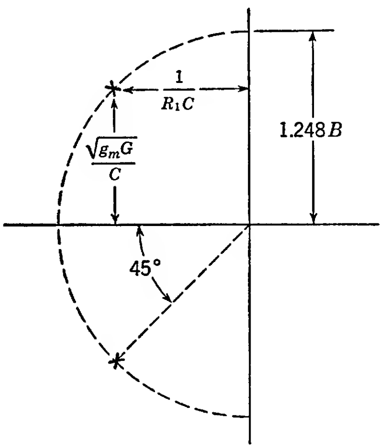


Fig. 10.27. Poles of the feedback pair.

Now let us consider two identical maximally flat feedback pairs. The poles of one pair are shown in Fig. 10.27, where G_1 is the plate load conductance, C is the interstage capacitance, and G is the feedback resistance. (It should be noted that the dimensions of Fig. 10.27 are subject to some approximations.) The over-all bandwidth is B and the plate loads G_1 and capacitances C are assumed the same for all four tubes. We find $R_1 = 1.135/BC = 1/(g_mG)^{1/2}$. The gain at $\omega = 0$ is not simply g_mR_1. Since the phasor length from each pole is $(2)^{1/2}$ times that when no feedback exists, the mean gain is $g_mR_1/(2)^{1/2} = 0.802g_m/BC$. This is $0.802/0.435 = 1.84$ times as large as the gain per R-C stage and hence the over-all gain is 11.5 times as large.

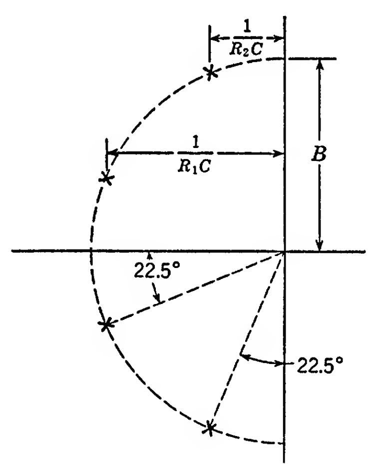


Fig. 10.28. The four-pole function using feedback pairs.

So far, we have talked of cascades of identical functions, with the cascade of shunt-peaked circuits and series-peaked circuits evidently most efficient. Now let us consider a single over-all n-pole maximally

flat function of bandwidth B. We can, for example, use two *different* feedback pairs to give poles as in Fig. 10.28, where subscripts 1 and 2 apply to the two different pairs respectively. We find $R_1 = 1/(BC \cos 22.5°)$ and $R_2 = 1/(BC \sin 22.5°)$. The gain per stage of the first pair at $\omega = 0$ is $g_m R_1 \cos 22.5° = g_m/BC$. Similarly, the gain of each tube in the second pair is $g_m R_2 \sin 22.2° = g_m/BC$. This fortuitous occur-

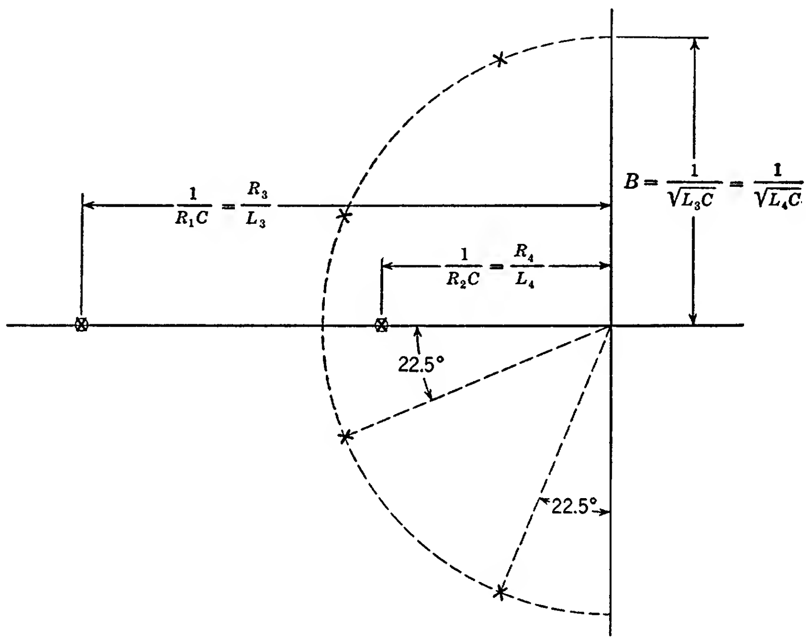


Fig. 10.29. The four-pole function with R-C and shunt-peaked circuits.

rence enables us to calculate the mean stage gain of a $2n$-tube maximally flat $2n$-pole function utilizing n feedback pairs simply as g_m/BC. The over-all gain advantage of this four-stage amplifier as compared to the four-stage R-C amplifier is then simply $(1/0.435)^4 = 28$.

Now let us study a system with two R-C stages and two shunt-peaked stages worked into a single four-pole maximally flat function. Let 1, 2, 3, 4 subscripts designate the four stages. The p-z plot is shown in Fig. 10.29. We find $L_3 = L_4 = 1/B^2C$, $R_1 = 2 \cos 22.5°/BC$, $R_2 = 2 \sin 22.5°/BC$, $R_3 = 1/(2BC \cos 22.5°)$, and $R_4 = 1/(2BC \sin 22.5°)$. The total gain at $\omega = 0$ is therefore $(g_m/BC)^4$. As with feedback pairs, we find that we may calculate the mean stage gain of a $2n$-tube maximally flat $2n$-pole function utilizing n pairs of R-C and shunt-peaked circuits

as simply g_m/BC. [Had we used an extra R-C stage here (or with feedback pairs) to give a single maximally flat function with an odd number of poles, the mean gain-bandwidth product would still be g_m/C.]

We could go on as we have done calculating various amplifiers and comparing their over-all gains to that of four stages of R-C circuits for the same over-all bandwidth. For example, we could use two shunt-peaked circuits and two series-peaked circuits (with the zeros of the shunt-peaked circuit cancelled) to get a very efficient over-all eight-pole gain function. Other combinations could be dreamed up as well. For example, we might try a cascade of four series-shunt-peaked circuits to give a highly efficient transfer function having both poles and zeros. However, enough examples have been given to demonstrate the design techniques adequately.

10.10 Some practical matters

Vacuum tubes of a given type will rarely be identical. Some have transconductances smaller than others (although still considered good tubes) which causes them to yield somewhat less gain than might be obtained with different tubes. In addition, plate resistances and distributed capacitances may vary. In fact, variations in tube parameters of 10 or 15 per cent from tube to tube are to be expected.

Circuits that are overly sensitive to the magnitude of tube capacitance and plate resistance, and feedback circuits sensitive to transconductance, may become mistuned or maladjusted if tubes are changed. Such circuits should be avoided where possible; otherwise, methods for making tuning or gain adjustments must be furnished.

The inclusion of means for making tuning adjustments is not always desirable or necessary in low-pass amplifiers, although their inclusion is often necessary in band-pass amplifiers. Capacitance can be varied by including small variable capacitors in the circuit in shunt with tube capacitances or elsewhere. However, if it is important that distributed shunt capacitances be kept to a minimum, variable capacitors should be avoided because of the additional circuit capacitance they introduce. If coils are present, they can be tuned by means of a slug inserted in the middle of the coil. A copper slug decreases inductance, whereas a ferrous slug increases inductance. Coils may be "trimmed" by adjusting the number and/or spacings of the turns. This method for making tuning adjustments avoids both slugs and variable capacitors.

When stages of an amplifier handle fairly large signals (which would normally be those stages near the output), some saturation may occur. The actual gain function of frequency may be quite different from that when the signal level is small. Saturation effects are common in ampli-

fiers tuned to give a maximally flat or similar type of transfer function with saturation occurring in the stage or stages furnishing the poles nearest the edge of the band. A typical result of saturation is to give a tilted shape to a response that would otherwise be flat. If such effects must be minimized, it is best to use simple one-pole functions for those stages of the amplifier which must handle fairly large signals.

Almost all amplifiers have incorporated in them a means for changing the gain. Three methods are generally employed. The most obvious is to use a variable resistor in the grid circuit of a tube as a manual gain control. This method is not desirable in wide-band amplifiers because of the large capacitance introduced by the potentiometer and because its variation will have some effect on the tuning of the circuit. In fact, trouble on this account is often met in audio-frequency circuits where the resistance load may be on the order of 1 megohm.

A second method for obtaining gain control, useful when pentode tubes are employed, is to supply one or more of the screens from a variable-voltage power supply, the variation being accomplished by means as simple as that afforded with a potentiometer. Reducing the screen voltage of a pentode reduces its transconductance and hence the gain. This method is very effective when applied to several tubes simultaneously. However, if the transconductance is changed too much, some of the parameters of the tube (capacitance and plate resistance) may be changed enough to cause detuning. In the case of tubes handling large signals, it must be remembered that a reduced transconductance may introduce saturation because the dynamic range is decreased.

The third method for obtaining gain control is that of changing the grid bias. The grid-bias voltage has a strong effect upon the transconductance (particularly with "remote cutoff" tubes). Bias control can be effected with a potentiometer as a by-passed cathode bias resistor where stray capacitances are unimportant. Alternately, the grid of a tube may be returned to an adjustable negative voltage. Bias variations permit a very large change in gain to be obtained, particularly if a negative voltage is applied to the grids of several tubes of a multitube amplifier simultaneously. With an "automatic volume control" (AVC) circuit (which is common in band-pass amplifiers but not in low-pass systems), the signal at the output of the amplifier is converted to direct current, filtered (so that the filtered signal has negligible components at frequencies falling in the pass band of the amplifier), and applied at such a polarity to the grids of the tubes that an increase in the output voltage tends to reduce the gain of the amplifier. The result is that the output of the amplifier will change but little in magnitude even though the input (carrier level) changes considerably.

Any gain control that changes the transconductance of a tube will also change the plate resistance in almost direct proportion because the amplification factor $g_m r_p$ remains approximately constant. Thus, circuits may be detuned considerably when the gain is changed because of the change in plate resistance. This is most noticeable in triodes but not so much in pentodes because the plate resistance of a pentode can more easily be made quite large compared to the impedance level of the interstage network with which the pentode operates.

10.11 Step-function response

The response of a network to a transient input disturbance is frequently of more direct interest than is its steady-state behavior. Although the steady-state characteristics and the transient response are intimately related (one can be calculated from the other), it is frequently more significant to calculate the transient response directly.

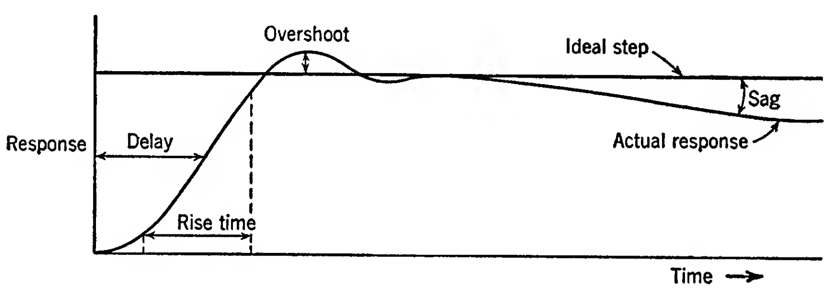


Fig. 10.30. Typical step-function response.

The standard input waveform used in transient calculations is the step function $EU(t)$, which is zero for $t < 0$ and constant at E for all $t > 0$; at $t = 0$, the input rises abruptly from zero to a constant E. The response of a system to a step function is usually adequate for determining the most important transient characteristics of the system. A typical low-pass step-function response (normalized to unity midband gain) is shown in Fig. 10.30. The ideal response can be considered to be the step function itself. The four most significant characteristics of the actual response are the "rise time," the "overshoot," the "delay," and the "sag."

The rise time is the time period required for the output to go between 10 and 90 per cent of its ideal final value. When the rise time is short, the amplifier is said to be "fast." Approximately, the rise time is inversely proportional to the bandwidth, and in low-pass systems is ordinarily dependent upon only the high-frequency p-z.

A factor of merit often used with wide-band amplifiers is the gain divided by the rise time, which is quite analogous to the gain-bandwidth product. For the simple R-C interstage network driven with a pentode, this factor is

$$\frac{\text{Gain}}{\text{Rise time}} = \frac{g_m R}{2.2RC} = \frac{g_m}{2.2C} \tag{10.20}$$

where C is the total interstage capacitance.

Approximately, the gain/rise time factor varies according to circuitry much as does the gain-bandwidth product. However, the factor can be made appreciably larger for a given bandwidth if some overshoot is permitted as compared to a system having the same bandwidth that does not introduce overshoot. The advantage obtained by employing four- rather than two-terminal interstage networks is appreciable, although the improvement in the gain/rise time factor may not be as large as the improvement in the gain-bandwidth product unless an appreciable overshoot can be tolerated.

When identical stages are cascaded and when the overshoot due to one stage is small (1 or 2 per cent or less), the rise time increases about as the square root of the number of cascaded stages. When nonidentical stages are cascaded, the over-all rise time can be calculated approximately as the square root of the sum of the squares of the individual rise times. Therefore, if one stage is appreciably slower than all the rest, the rise time will be largely dependent upon this one stage. When the overshoot due to one stage is larger (5 to 10 per cent), the over-all rise time will be appreciably less than that calculated from the square root relation.

A useful rule to remember is that when overshoot is small the product of the amplifier over-all half-power (radian) bandwidth and the 10–90-per-cent rise time (in seconds) lies in the range 2.2 to 2.8. When the overshoot is appreciable, this factor will be large, while values around 2.2 are applicable when the overshoot is small.

Previously, we compared various amplifier functions to the simple R-C interstage network in terms of the improvement in the gain-bandwidth product. We can make a similar comparison on the basis of the gain/rise time factor. The ratio of the gain/rise time for any circuit to that for the simple R-C circuit is termed the "relative speed" of the circuit.

The overshoot is, as the term implies, the amount by which the response exceeds the ideal value. It is usually expressed as a percentage and, like the rise time, is dependent upon the high-frequency p-z. When considerable overshoot exists, the output tends to ripple somewhat about the ideal value before finally settling down.

Overshoot is a direct result of oscillatory transient partial outputs and consequently will not occur if all the poles of the transfer function are purely real (except in special cases such as an overpeaked cathode-peaking circuit). When a transfer function does have complex poles, the severity of the overshoot is more or less proportional to the closeness of the poles to the $j\omega$ axis. For this reason, Chebyshev functions are normally not employed for amplifiers designed to handle transient disturbances. Rather, linear-phase functions are preferred with maximally flat functions utilized when high speed rather than overshoot is of primary importance. Large improvements in the gain/rise time ratio can be realized by admitting even small amounts of overshoot.

When several stages are cascaded, the overshoot of the cascaded system may or may not be larger than the overshoot attributable to one stage. As stages having small overshoot (1 or 2 per cent) are cascaded, the overshoot increases slowly if at all. When the individual stages have overshoots on the order of 5 or 10 per cent, the overshoot increases about as the square root of the number of stages cascaded.

Some representative figures of overshoot are worthwhile additions here. All-pole maximally flat functions with one, two, three, four, and six poles have overshoots of 0, 4.3, 8.15, 10.9, and 14.3 per cent respectively.

A cascade of identical one-pole functions (which never have overshoot) has a τB of 2.2 for all numbers of cascaded stages, where B is the over-all bandwidth.

A cascade of one, two, three, four, and six identical two-pole maximally flat functions has an overshoot of 4.3, 6.25, 7.7, 8.4, and 10 per cent respectively.

A cascade of one, two, and four identical three-pole maximally flat functions has an overshoot of 8.15, 11.2, and 14.2 per cent respectively.

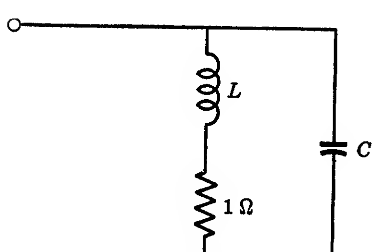


Fig. 10.31. The shunt-peaked circuit.

The two-pole Q-D function (see Chap. 5) is also the two-pole maximally flat function. A three-pole Q-D function has an overshoot of about 12 per cent.

In addition to the previous figures, some data regarding specific networks are of value. The shunt-peaked circuit of Fig. 10.31 can be described with the parameter L/C (for 1 Ω resistance). There is no overshoot for $L/C \leq 0.25$. The circuit is maximally flat when $L/C = (2)^{1/2} - 1$. For $L/C = 0$, 0.25, 0.414, 0.5, and 0.6, the overshoot is 0,

0, 3.1, 6.7, and 11.4 per cent respectively, and the relative speed is 1.0, 1.4, 1.7, 1.9, and 2.1 respectively.

The linear-phase circuit with element values as given in Fig. 10.32 has a speed advantage of 1.77 and an overshoot of only 1 per cent.

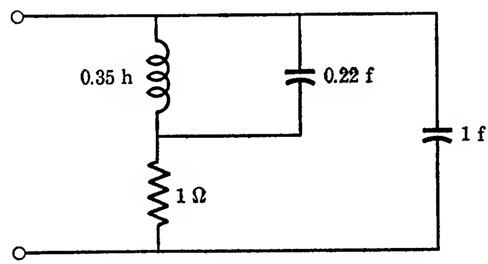


Fig. 10.32. The linear-phase circuit.

A four-terminal linear-phase type of network is shown in Fig. 10.33. It has a relative speed advantage of about 2.5 with only 1-per-cent overshoot. A 2:1 capacitance ratio is employed.

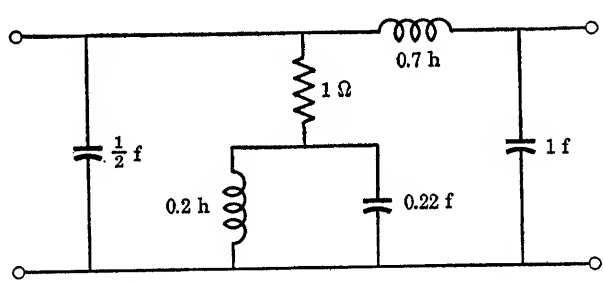


Fig. 10.33. The series-shunt-peaked circuit.

The delay of the output step function is another measure of transient response. It is the time period by which the actual output lags the ideal step-function output and is measured from $t = 0$ to approximately the 50-per-cent value on the rising part of the actual response. The delay was discussed in Chap. 5 in terms of Q-D functions and was shown to be given by

$$\text{Delay} = -\Sigma \text{ pole positions} + \Sigma \text{ zero positions} = T - \tau \qquad (10.21)$$

where the pole positions are in themselves negative and low-frequency p-z are ignored. If the transfer function $F(p)$ is expressed in polynomial form (considering only high-frequency p-z), the delay can be determined

from

$$F(p) = H\frac{1 + \tau p + a_2p^2 + \cdots}{1 + Tp + b_2p^2 + \cdots} \tag{10.22}$$

The sag is, as its name implies, the amount by which the output falls below the ideal output. It is usually expressed as a percentage of the ideal output over a prescribed time interval, for example, 10-per-cent sag in 100 microseconds. Sag is due entirely to the low-frequency p-z and is a consequence of the inability of an a-c amplifier to support a d-c output.

Most low-pass amplifiers have low- and high-frequency effects that are widely separated in frequency. Therefore, the low- and high-frequency transient responses can generally be calculated separately and then superimposed, the low-frequency sagging behavior being pieced on the end of the fast-rising part of the response.

If all the high-frequency p-z are neglected, the transfer function gives the low-frequency behavior. This function must have an equal number of p-z; if it does not, the gain will not be constant at mid-band frequencies. Thus, the low-frequency transfer function will have the form

$$F(p) = \frac{p^n + a_{n-1}p^{n-1} + a_{n-2}p^{n-2} + \cdots + a_{n-q}p^{n-q}}{p^n + b_{n-1}p^{n-1} + \cdots + b_1p + b_0} \tag{10.23}$$

Let a step function be applied at the input of the system at $t = 0$. A very short time after $t = 0$ all the integrals of the input must be very small. Therefore, we are led to divide the numerator and the denominator of eq. 10.23 by p^n and expand the denominator into the numerator as

$$\begin{aligned} F(p) &= \frac{1 + (a_{n-1}/p) + (a_{n-2}/p^2) + \cdots}{1 + (b_{n-1}/p) + (b_{n-2}/p^2) + \cdots} \\ &= 1 + \frac{a_{n-1} - b_{n-1}}{p} + \frac{a_{n-2} - b_{n-2} - a_{n-1}b_{n-1} + b_{n-1}{}^2}{p^2} + \cdots \end{aligned} \tag{10.24}$$

For small time, the response to the unit step function can therefore be written

$$\begin{aligned} e(t) = U(t) &+ (a_{n-1} - b_{n-1})\int U(t)\,dt \\ &+ (a_{n-2} - b_{n-2} - a_{n-1}b_{n-1} + b_{n-1}{}^2)\iint U(t)\,dt\,dt \\ &+ \text{higher order integrals} \end{aligned} \tag{10.25}$$

where we have interpreted p to be the derivative operator. Integrating eq. 10.25 from $t = 0$ to time t, we get

$$e(t) = U(t) + (a_{n-1} - b_{n-1})t + (a_{n-2} - b_{n-2} - a_{n-1}b_{n-1} - b_{n-1}{}^2)\frac{t^2}{2} + \text{higher order powers of time} \quad (10.26)$$

The initial slope of the output is simply $a_{n-1} - b_{n-1}$. Since a_{n-1} is the negative of the sum of the zero positions, and b_{n-1} is the negative of the sum of the pole positions, and because all p-z normally have negative real parts, the slope will be downward if the zeros are nearer the origin than the poles (not counting the zeros exactly at the origin), and conversely. The initial slope can be made zero with a low-frequency compensating network. Then, the output is

$$e(t) = U(t) + (a_{n-2} - b_{n-2})\frac{t^2}{2} + \text{higher powers of time} \quad (10.27)$$

The long-time step-function behavior is shown in typical form for various amounts of compensation in Fig. 10.34. A considerable im-

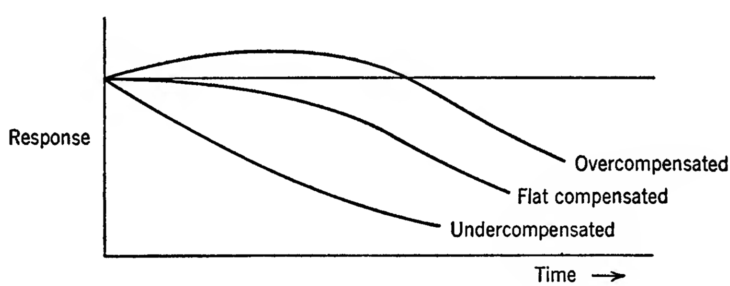


Fig. 10.34. Long-time step-function response.

provement in the long-time behavior results from nearly flat compensation.

A practical amplifier is never driven with an ideal step function. Rather, the rise time of a practical input step function is best made comparable to the rise time of the response of the amplifier to a perfect step function. There is no point in overdesigning a transient amplifier. In other words, there is no justification for building an amplifier very much faster than the expected input signal. This is equivalent to saying that it is pointless to build an amplifier with a bandwidth much

larger than the bandwidth of the signal source used to drive the amplifier.

It is often helpful to hypothesize a fictional amplifier which, when driven with an ideal step function, yields the practical amplifier input. The p-z of the practical amplifier, along with those of the hypothetical amplifier, can then be worked into a single over-all function. When this is done, the practical amplifier itself may have a rather poor response to

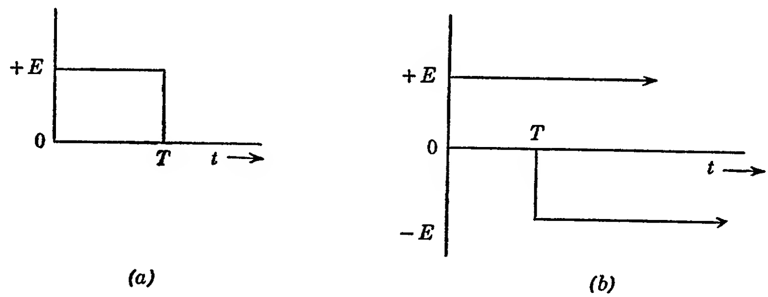


Fig. 10.35. Construction of an ideal pulse.

an ideal step function, although the response to the practical input step function may be quite satisfactory.

Input signals other than the step function are often of interest, an especially important one being the rectangular pulse diagrammed in Fig. 10.35*a*. If it is observed that a pulse of duration T can be constructed from a step function at $t = 0$ followed by a negative step function at $t = T$, then the method for evaluating the pulse response from

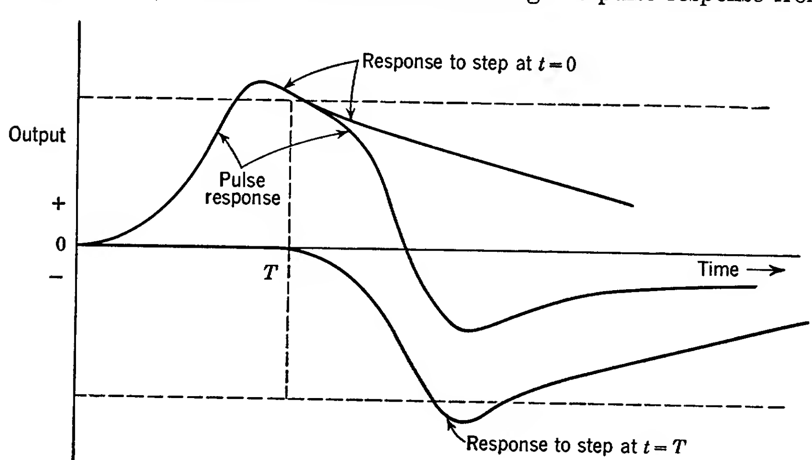


Fig. 10.36. Calculation of pulse response.

the step-function response should be clear. All that must be done is to add to the step-function response at $t = 0$ the response due to the negative step at $t = T$. This is indicated in Fig. 10.36. In a similar manner, the response to a more arbitrary input can be constructed by adding together the responses to several step functions of various amplitudes and times of occurrence, as directed from the approximation to the actual input with a number of superimposed steps.

Problems

1. The amplifier of Fig. P.1 is to have a two-pole maximally flat transfer function e_3/e_1 with a bandwidth of 8 mcs. Determine R_1, L, and R_2 assuming $g_m = 5000$ micromhos, $r_p = 1\text{M}$ and 20K, $C_1 = 17\ \mu\mu\text{f}$, and $C_2 = 200\ \mu\mu\text{f}$. Determine the voltage gain at $\omega = 0$. Also determine the power gain at $\omega = 0$ if the grid-circuit impedance is 500K. What peak output signal can be furnished by the cathode follower if $i_b = 30$ ma at the operating point?

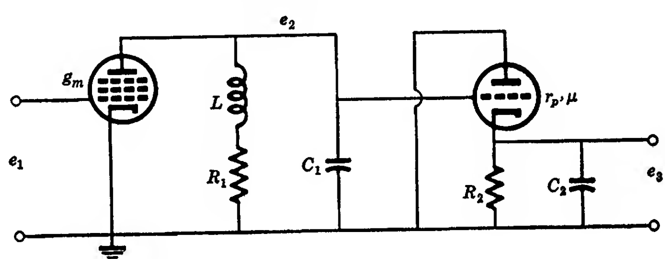


Fig. P.1.

2. Repeat Prob. 1 using a two-pole equal-ripple function with $1/(1 + \epsilon)^{1/2} = 0.9$ and having the same half-power bandwidth.

3. An audio-frequency voltage amplifier using a cascade of three triodes with $\mu = 20$ and $r_p = 50\text{K}$ is to have an over-all bandwidth of 10 kcs. The capacitance C is 20 $\mu\mu$f. Determine R such that the voltage gain e_3/e_0 at $\omega = 0$ is an approximate maximum. Include Miller capacitance in your calculations. Use $C_{gp} = 5\ \mu\mu\text{f}$. Refer to Fig. P.3.

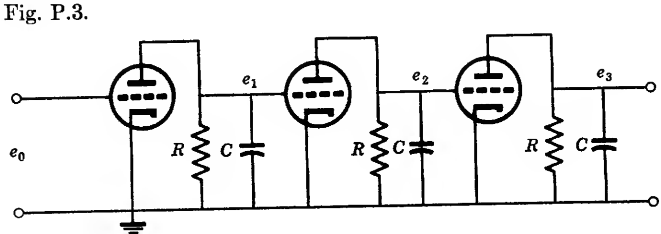


Fig. P.3.

4. An audio volume control has the equivalent circuit shown in Fig. P.4. Calculate e_2/e_1 and plot the gain characteristics when the arm of the control is $\frac{1}{3}$, $\frac{2}{3}$, and all the way up from the ground side.

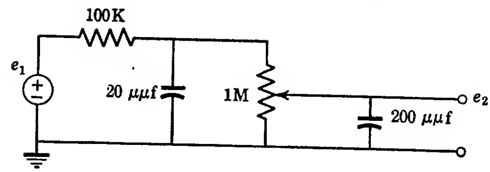


Fig. P.4.

5. Derive the design equations for cathode peaking; that is, derive the proper relations between the resistances and capacitances of Fig. P.5 in order that p-z cancellation take place.

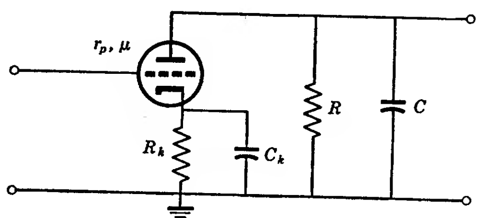


Fig. P.5.

6. Cathode peaking may also be used with the shunt-peaked circuit as indicated in Fig. P.6. Derive the design relations for the maximally flat transfer function.

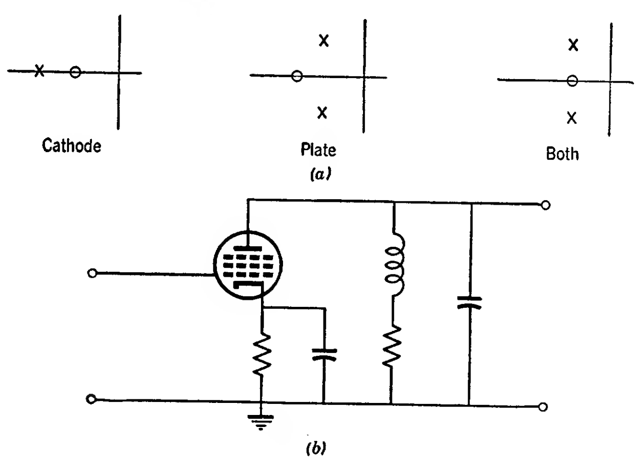


Fig. P.6.

7. It is desired to build a two-terminal low-pass interstage having a frequency above the pass band where the transfer function goes to zero. The plate network and the p-z of the input impedance are shown in Fig. P.7. For an amplifier using three (approximately) maximally flat stages of this sort with a bandwidth of 10 mcs and with zeros at 20 mcs, determine the gain at $\omega = 0$ and sketch the magnitude of the over-all transfer function. Determine the network element values.

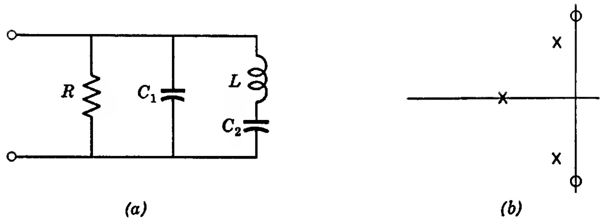


Fig. P.7.

8. One method of feeding voltage to the plate of a tube is through a large inductance as shown in Fig. P.8, which is termed "choke feed." Determine the p-z of the transfer function assuming C_c is very large. Derive an expression for the low-frequency cutoff point and also derive an expression which must be obeyed if the p-z of the transfer function are to lie on the negative real axis.

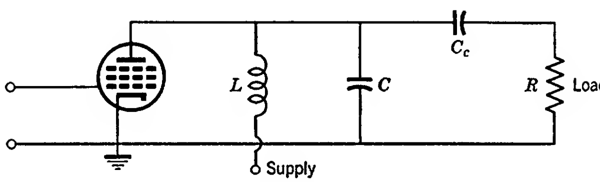


Fig. P.8.

9. An amplifier has N identical stages. Derive an expression for the low-frequency half-power point resulting from R-C coupling networks.

10. An amplifier has N identical stages. Derive an expression for the low-frequency half-power point resulting from cathode resistors and by-pass capacitors.

11. An amplifier has N identical stages. Derive an expression for the low-frequency half-power point which results from screen resistors and by-pass capacitors.

12. Design a compensating network to completely cancel cathode by-pass effects.

13. Design a compensating network to completely cancel screen by-pass effects.

14. A given stage has cathode and screen by-passing and a coupling network. Derive the equation from which a compensating circuit can be designed to compensate for the screen, cathode, and coupling all at the same time.

15. Is it more expensive to obtain by-passing at the screen or at the cathode? (Expense is proportional to the product of the capacitance and the d-c working voltage of the by-passing capacitor.)

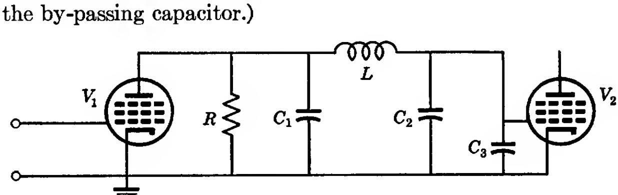


Fig. P.16.

16. A two-terminal-pair interstage having a maximally flat three-pole transfer function is shown in Fig. P.16. C_1 is the output capacitance of V_1 and C_2 is the

input capacitance of V_2. C_2 must be padded with C_3 (unless C_2 happens to be the proper value to use) so that $C_2 + C_3 = 3C_1$ in order that the transfer function be maximally flat. What is the advantage factor compared to the R-C interstage having a total interstage capacitance of $C_1 + C_2$? Assume $C_2 = 2C_1$.

17. If the network of Prob. 16 is turned end for end, the circuit of Fig. P.17 results, where now it is necessary to add a capacitance C_3' in parallel with C_1 such that $C_1 + C_3' = 3C_2$. What is the advantage factor as compared to the R-C interstage having a total interstage capacitance of $C_1 + C_2$? C_1 and C_2 are the same as in Prob. 16.

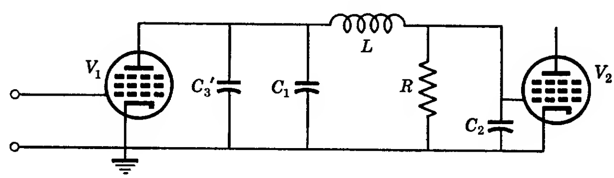


Fig. P.17.

18. Repeat Prob. 16 using the Q-D function instead of the maximally flat function.

19. A typical audio output stage is shown in Fig. P.19. Normally, the coefficient of coupling of the transformer is close to unity. Locate the p-z of the transfer function and sketch the gain as a function of frequency.

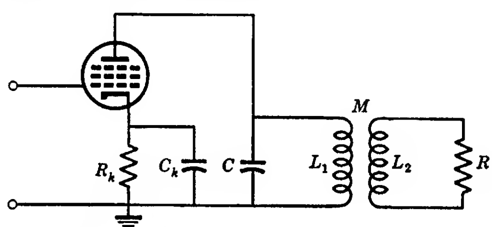


Fig. P.19.

20. Determine suitable values of L_1, L_2, C, and C_k for the circuit of Fig. P.19 if $R = 5\ \Omega$, $R_k = 300\ \Omega$, $L_1/L_2 = 100$, and $k = 0.98$ in order that the gain characteristics be reasonably flat and have lower and upper cutoff points of 120 and 8000 cps respectively.

21. It is desired to build an amplifier having a bandwidth of 4.8 mcs and a voltage gain of 10,000. If it can be done, how many R-C stages are required using tubes with a g_m/C of 50 mcs and what is the actual gain?

22. Repeat Prob. 21 using a cascade of identical shunt-peaked two-terminal interstage networks adjusted to be maximally flat.

23. Repeat Prob. 21 using a cascade of identical two-pole maximally flat functions with each two-pole function realized with a shunt-peaked circuit and an R-C circuit.

24. Repeat Prob. 21 using a cascade of identical feedback pairs. Use the maximally flat two-pole function for each pair.

25. Repeat Prob. 21 using a cascade of circuits such as that of Fig. P.16.

26. Repeat Prob. 21 using feedback pairs to give a single n-pole maximally flat function.

27. Repeat Prob. 21 using shunt-peaked and R-C interstage networks to get a single n-pole maximally flat function.

28. Adjust the parameters of the four-terminal interstage network of Fig. P.28 to give the maximally flat function and compare the bandwidth with that when $C' = 0$. Assume C' is an adjustable parameter for the purpose of setting the frequency of infinite attenuation. Can higher gain-bandwidth products be obtained with all-pole functions or with functions having zeros on the $j\omega$ axis outside the pass band (with the same number of poles)?

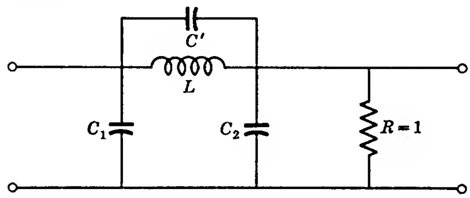


Fig. P.28.

29. An amplifier has N identical low-pass R-C (one-pole) interstage networks. Let each stage have a normalized bandwidth of unity. Calculate and plot the impulse response of the amplifier for $N = 1$, 2, and 4. Also plot the step-function response.

30. A transfer function is defined by two poles at p_1, $p_1^* = -1 \pm j\beta$. Calculate and plot the step-function response for $\beta = 0$, 0.5, 1, and 2. Is the real part of the pole position usually related to an R-C product?

31. Sketch the magnitude of the functions of frequency of Prob. 30. Discuss the relationship between frequency and transient response.

32. A two-pole maximally flat function has a bandwidth B. A rectangular pulse is applied. If the final value of the step-function response is unity, what input pulse width produces a pulse output that barely rises to unity? (Use graphical methods.)

33. Sketch the pulse response for the particular pulse width found in Prob. 32. Also sketch the response for pulses half this long and twice this long.

34. The impulse response of an amplifier achieves a height V_m. A very short rectangular pulse can approximate an impulse. If the height of the applied pulse is maintained constant, how does the height V_m vary with the width of the applied pulse T? (Short pulses are sometimes used for determining the impulse response experimentally.)

35. The step-function response available at the input to an amplifier is like the output of an hypothetical amplifier having the transfer function $[B/(p + B)]^2$ when driven with an ideal step. This signal is to be amplified by two shunt-peaked stages (δ not necessarily 1.554). Design the amplifier so that its output will be the same as the ideal step-function response of a transfer function $[5B/(p + 5B)]^4$. How does the rise time of the actual input compare with that of the actual output? Sketch the gain function of frequency of one of the amplifier stages.

11

Band-Pass Amplifiers

There is little essential difference between band-pass and low-pass amplifiers that are dependent upon two- and four-terminal interstage networks. Of course, the networks themselves will be somewhat different: they will be band-pass rather than low-pass networks. Thus, part of what we have to say here can be rather brief.

However, there are several types of band-pass amplifiers that do not have obvious low-pass counterparts, for example, devices that utilize feedback. In addition, there are certain band-pass systems that are rarely used for low-pass systems, for example, the cascode amplifier. We shall discuss typical circuits of this general type in sufficient detail to indicate the principal techniques of value.

11.1 Isolated stages with two-terminal networks

There is little need to go into a lengthy discussion regarding isolated amplifiers (grounded-cathode tubes) having two-terminal interstage networks. The principles are essentially identical to those of low-pass amplifiers, including bandwidth-narrowing considerations. Besides, there are only two important types, the "single-tuned" circuit and the "dominant-pole" circuit.

A reasonably simple two-terminal impedance can do little more than place a pair of complex-conjugate poles near the center frequency of the amplifier and fairly close to the $j\omega$ axis. Figure 11.1 shows all the essentials of the simplest band-pass functions. Figure 11.1*a* corresponds to the single-tuned circuit of Fig. 11.2, whereas Fig. 11.1*b* is that of the dominant-pole circuit in which one pair of complex-conjugate poles out of some arbitrary number of p-z largely determines the band-pass nature of the function.

The circuit of Fig. 11.2 can be driven with either a triode or a pentode. However, at all but the lowest frequencies, pentodes are generally employed because the Miller capacitance associated with triodes can be severe enough to cause instability if not carefully neutralized. In Fig. 11.2 G is the sum of the circuit conductance G' and the plate conductance

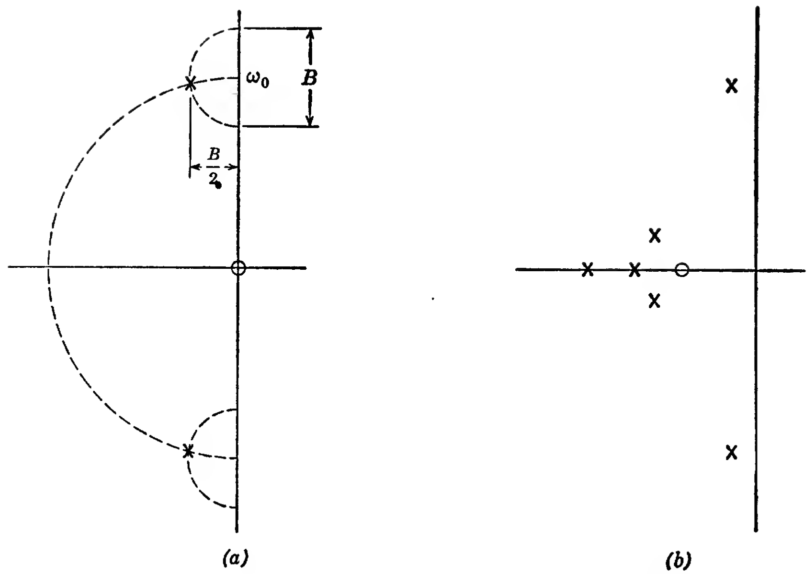


Fig. 11.1. Parallel-resonant and dominant-pole characteristics.

$1/r_p$. The transfer function is

$$\frac{e_1}{e_0} = -\frac{g_m}{C}\left(\frac{p}{p^2 + p(G/C) + (1/LC)}\right) \tag{11.1}$$

The poles are located at

$$p_1, p_1{}^* = -\frac{G}{2C} \pm j\left[\frac{1}{LC} - \left(\frac{G}{2C}\right)^2\right]^{1/2} \tag{11.2}$$

The bandwidth B can be identified in Fig. 11.1a as

$$B = G/C \tag{11.3}$$

which is twice the real part of the pole position.

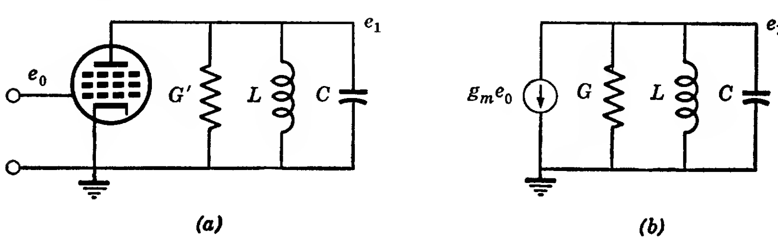


Fig. 11.2. Parallel-resonant circuit as a plate load.

The stage gain (at center frequency) times the bandwidth between half-power points is g_m/C, which is the same as that for the simple R-C low-pass function. Bandwidth narrowing is also the same. It should be noted that the bandwidth is the total bandwidth between half-power points, whereas for the R-C circuit it is the bandwidth measured from $\omega = 0$.

The center frequency is that where the phase shift is zero and also is the frequency of the peak of the amplitude characteristic. (The frequencies of zero phase shift and of the peak do not correspond exactly for most dominant-pole circuits except when the bandwidth is very small.) The center frequency of the single-tuned circuit is $1/(LC)^{1/2}$, which is the point of intersection of the $j\omega$ axis and the circle centered at the origin drawn through the poles.

The similarities between the single-tuned circuit and the low-pass R-C circuit should be noted by the reader. They result from the low-pass and band-pass equivalence of the two circuits.

The quality factor Q of a tuned circuit is an important and useful measure. For the single-tuned circuit where the resistance is in shunt with the inductance, it is

$$Q = \frac{\text{Center frequency}}{\text{Bandwidth}} = \frac{C/G}{(LC)^{1/2}} = \frac{R}{(L/C)^{1/2}} = \omega_0 RC = R/\omega_0 L \qquad (11.4)$$

The higher the Q, the more "selective" is the circuit.

The fundamental definition of Q is 2π times the energy stored in the circuit divided by the energy dissipated per cycle. It is easy to prove that this general definition leads to the same result as eq. 11.4 for the simple parallel-resonant circuit. If e_m is the peak value of the sine-wave voltage across capacitor C at resonance, then $e_m{}^2C/2$ is the energy stored in the parallel-resonant circuit. (When the energy in the capacitance is a maximum, that in the inductance is zero, and conversely.) The average power dissipated is clearly $e_m{}^2/2R$; hence, the energy dissipated *per cycle* is $e_m{}^2/2Rf_0$, where f_0 is the resonant frequency in cycles per second. Therefore, $Q = 2\pi(e_m{}^2C/2)/(e_m{}^2/2Rf_0) = \omega_0 RC$, which is eq. 11.4.

It is a simple matter to cascade single-tuned circuits. If all the circuits are identical, the resulting amplifier is said to be "synchronously tuned." Bandwidth narrowing is the greatest disadvantage of the synchronous-tuned amplifier. By making G and L (and perhaps C) different for different stages, single-tuned circuits can be used to build up any number of more complex all-pole band-pass functions, such as maximally flat, equal-ripple, linear-phase, and so on. A very common procedure is

to use a cascade of two- or three-pole maximally flat functions. For example, a nine-stage amplifier can be built with a cascade of three triples, each of which is identical and maximally flat.

Tuning an amplifier consisting of single-tuned circuits is particularly simple. A signal generator at the frequency of a pole is applied to the input and the interstage related to that pole is tuned (usually the inductance) to give a maximum reading on a voltmeter at the output. The same procedure is carried out for each pole. No matter how complex is the all-pole function and whether wide or narrow band, this procedure is exact for the single-tuned circuit. It is not exact for other dominant-pole band-pass circuits, although it becomes accurate in the narrow-band case.

11.2 Isolated stages with four-terminal networks

Only two four-terminal interstage networks are of widespread importance: the transformer-coupled interstage and the double-tuned capacitance-coupled interstage.

The transformer-coupled network is shown in Fig. 11.3 along with a plot of the typical positions of its p-z. The transformer network having two node pairs can be solved to obtain

$$\frac{e_2}{e_0} = -\frac{g_m k\omega_1\omega_2}{(1-k^2)(C_1C_2)^{1/2}} \cdot \frac{p}{p^4 + b_3p^3 + b_2p^2 + b_1p + b_0} \tag{11.5}$$

where

$$\begin{aligned} b_3 &= \frac{\omega_1}{Q_1} + \frac{\omega_2}{Q_2} \\ b_2 &= \frac{\omega_1\omega_2}{Q_1Q_2} + \frac{\omega_1^2 + \omega_2^2}{1-k^2} \\ b_1 &= \frac{\omega_1^2\omega_2}{Q_2(1-k^2)} + \frac{\omega_2^2\omega_1}{Q_1(1-k^2)} \\ b_0 &= \frac{\omega_1^2\omega_2^2}{1-k^2} \end{aligned} \tag{11.6}$$

and where

$$\begin{aligned} \omega_1 &= 1/(L_1C_1)^{1/2} \qquad & \omega_2 &= 1/(L_2C_2)^{1/2} \\ Q_1 &= 1/(\omega_1L_1G_1) \qquad & Q_2 &= 1/(\omega_2L_2G_2) \end{aligned} \tag{11.7}$$

The denominator of eq. 11.5 cannot be factored in general. Such networks must be synthesized through either approximation or coeffi-

cient matching. In coefficient matching, a desired polynomial is set up and the coefficients b_0, b_1, b_2, and b_3 are equated to the numerical values of the coefficients of the known polynomial. This method can always be employed providing the network is capable of having p-z in the desired positions. However, the numerical problems are severe if the bandwidth is small (that is, small differences between large numbers).

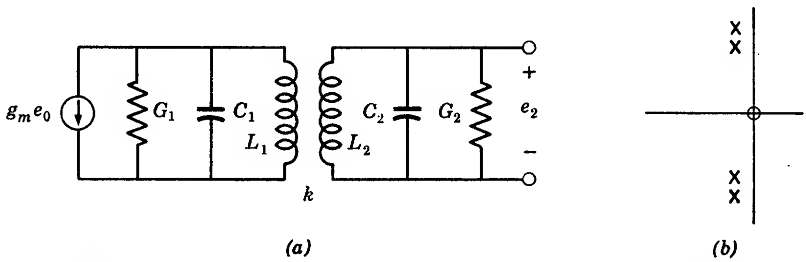


Fig. 11.3. Equivalent circuit of the transformer-coupled interstage.

Rather than solve this general case, let us obtain approximations of importance for narrow-band functions. If $Q_1 = Q_2 = Q$, $\omega_1 = \omega_2 = \omega_0$, and $k^2 \ll 1$ in eq. 11.5, approximate pole positions can be found as

$$p_1, p_1^*, p_2, p_2^* = -\frac{\omega_0}{2Q} \pm j\omega_0\left(1 \pm \frac{k}{2}\right) \tag{11.8}$$

For this "equal-Q" condition, it is an easy matter to find the desired circuit parameters for a maximally flat or any other pole configuration requiring two pairs of complex-conjugate poles where the poles are the same distance from the $j\omega$ axis. It should be apparent from eq. 11.8 that, for a maximally flat function, it is necessary that $k = 1/Q$. This particular value of k is called "critical coupling." When critical coupling is employed, the bandwidth is $B = (2)^{1/2}\omega_0 k = (2)^{1/2}\omega_0/Q$. The mid-band gain is most easily determined from the pole positions and the multiplier of eq. 11.5. The phasor length from the zero at the origin to the center frequency is simply ω_0. Each pole near $-j\omega_0$ has a phasor length of very nearly $2\omega_0$. Each pole near $+j\omega_0$ has a phasor length $B/2$. Thus, the mid-band gain is approximately $\omega_0/[(2\omega_0)^2(B/2)^2]$ times the multiplier of eq. 11.5, which (assuming $k^2 \ll 1$) is $g_m k\omega_0/(C_1C_2)^{1/2}$. From these various relations, the gain-bandwidth product is calculated to be

$$(GB)_{\text{stage}} = g_m/(2C_1C_2)^{1/2} \tag{11.9}$$

which can be as large as $(2)^{1/2}$ times that of the single-tuned circuit if C_1 and C_2 are each half the total interstage capacitance.

Another special case of interest is that when one of the Q's is much larger than the other. Then, one of the two conductances can be approximated as zero. Assume $G_1 = 0$ so that $Q_1 = \infty$. Also, assume that $\omega_1 = \omega_2 = \omega_0$ and $k^2 \ll 1$. Then, we can find approximate pole positions as

$$p_1, p_1^*, p_2, p_2^* = -\frac{\omega_0}{4Q_2} \pm j\omega_0 \left[1 \pm \frac{[k^2 - (1/2Q_2)^2]^{1/2}}{2}\right] \tag{11.10}$$

If $Q_2 = \infty$ instead of Q_1, eq. 11.10 will apply if Q_2 is replaced with Q_1.

The poles of eq. 11.10 will have the same real part only if $k > 1/2Q_2$. For values of k less than this, the poles lie along a line parallel to the real axis of the p plane. When $k = 1/2Q_2$, the poles lie at the same point.

The poles for one Q infinite can easily be placed to give a maximally flat or similar function. For the maximally flat function, the bandwidth is $\omega_0/(2^{1/2}Q_2)$, which requires that $k = 1/(2^{1/2}Q_2)$. The gain-bandwidth product is

$$(GB)_{\text{stage}} = g_m/(C_1C_2)^{1/2} \tag{11.11}$$

which can be as much as twice that of the single-tuned circuit.

When identical maximally flat transformer-coupled stages are cascaded, bandwidth narrowing is the same as that for cascaded low-pass two-pole maximally flat functions; for N stages it is $(2^{1/N} - 1)^{1/4}$. Multiplying the narrowing factor by the stage gain-bandwidth product gives the system gain-bandwidth product.

Of course, several transformer-coupled stages can be cascaded with no two being the same to give a single over-all maximally flat, equal-ripple, linear-phase, or similar function, having an even number of poles. (An extra single-tuned circuit can be used to make the over-all function have an odd number of poles.) Such systems have been called "stagger damped" because the most observable physical difference between the various interstages is the difference in the loading resistors.

The narrow-band approximation in which the poles are placed on a semicircle is a good approximation for the double-tuned mutually coupled circuit up to bandwidth-to-center-frequency ratios of about two. It is rare that bandwidths larger than this are required. (See Chap. 4.)

Another frequently employed mutually coupled interstage network is that shown in Fig. 11.4a. This circuit is of most value in tuned radio-frequency amplifiers where a small bandwidth is suitable and when large gain-bandwidth products are not of primary importance. The circuit is likely to be found in all but the cheapest communications receivers. Only one element need be changed, usually C, in order to tune the network. Several stages may be tuned simultaneously by mechani-

cally ganging their capacitors together. Normally, the transformer is made to have an appreciable step-up ratio and the coefficient of coupling is made small. The advantage of a large step-up ratio is that the over-all voltage gain can be made quite large at resonance without the voltage gain from grid to plate being so large that stray feedback becomes a problem.

The equivalent circuit of Fig. 11.4*a* is shown in Fig. 11.4*b*. Let us assume that the primary impedance and coupling coefficient are so small

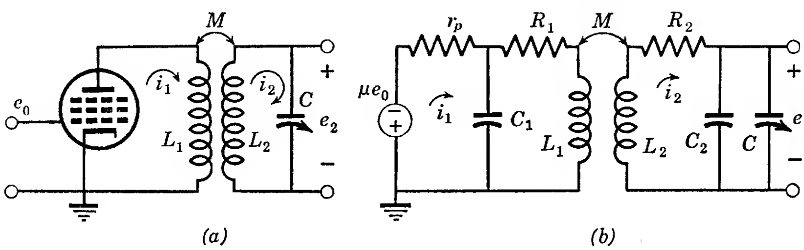


Fig. 11.4. The single-tuned mutually coupled circuit.

that the primary stray capacitance C_1 can be ignored. The secondary stray capacitance C_2 will be assumed to be part of capacitance C. Also, the primary wire resistance will be neglected in comparison with the plate resistance. These assumptions permit us to calculate the primary current without regard for the secondary circuit as

$$i_1 = \frac{-\mu e_0}{r_p + pL_1} \tag{11.12}$$

The voltage induced in the secondary as a result of the primary current is $\pm Mpi_1$. The current that flows in the secondary is therefore

$$i_2 = \frac{\pm Mpi_1}{R_2 + pL_2 + 1/pC} \tag{11.13}$$

Combining eqs. 11.12 and 11.13 (noting also that $e_2 = i_2/pC$), we get

$$\frac{e_2}{e_0} = \frac{\mp \mu Mp}{(p^2L_2C + pR_2C + 1)(pL_1 + r_p)} \tag{11.14}$$

The single real pole at $p = -r_p/L_1$ is very far from the origin when a pentode is employed because r_p is large. Thus, this pole may be neglected; that is, we can ignore pL_1 in the factor $(pL_1 + r_p)$. Then, sub-

stituting $\mu = g_m r_p$, $\omega_0{}^2 = 1/L_2C$, and $Q = \omega_0 L_2/R_2 = 1/\omega_0 R_2 C$, we get

$$\frac{e_2}{e_0} = \frac{\mp g_m M \omega_0{}^2 p}{p^2 + p\omega_0/Q + \omega_0{}^2} \tag{11.15}$$

which is functionally the same as the transfer function of a pentode with a parallel-resonant two-terminal load impedance. The gain at band center is $g_m M \omega_0 Q$ and the bandwidth is ω_0/Q.

When the circuit is tuned, capacitance C is normally varied. As C increases, frequency ω_0 decreases as $(1/C)^{1/2}$ and Q decreases (approximately) as $(1/C)^{1/2}$. Thus the gain at band center decreases as $1/C$ and

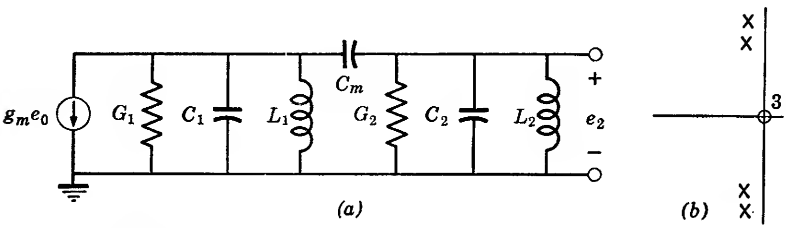


Fig. 11.5. The double-tuned capacitance-coupled circuit.

the bandwidth remains approximately constant. These effects limit the practical frequency tuning range to something on the order of 2:1 or 3:1. A greater tuning range with approximately constant gain can be obtained by varying the inductance rather than the capacitance or by varying both (as is done in "butterfly" resonant circuits).

The double-tuned capacitance-coupled circuit also makes use of two resonant circuits but, unlike transformer-coupled circuits, coupling is by means of a capacitor rather than a mutual inductance. Its circuit and typical p-z locations are shown in Fig. 11.5. There are three zeros at the origin rather than one as with transformer coupling. The reader can observe that this has little effect when the bandwidth is small but causes the amplitude function to become lopsided in the wide-band case with the voltage gain falling off only as $1/\omega$ at frequencies far above those of the poles. One result of the three zeros at the origin is that the gain-bandwidth product is no better than that of the single-tuned circuit for large bandwidths. However, when the bandwidth is small, the gain-bandwidth product is as large as that of the transformer-coupled circuit. We shall study only the narrow-band network here, which occurs for C_m small compared to C_1 and C_2.

Consider first the equal-Q situation in which both resonant circuits are identical; that is, $C_1 = C_2 = C$, $L_1 = L_2 = L$, and $G_1 = G_2 = G$.

The coefficient of coupling can be defined as $k = C_m/C$. The transfer function is

$$\frac{e_2}{e_0} = -\frac{g_m k}{C(1+2k)} \frac{p^3}{p^4 + b_3p^3 + b_2p^2 + b_1p + b_0} \tag{11.16}$$

where

$$\begin{aligned} b_3 &= \frac{2\omega_0(1+k)}{Q(1+2k)} \\ b_2 &= \frac{2\omega_0^2(1+k)}{1+2k} + \frac{\omega_0^2}{Q^2(1+2k)} \\ b_1 &= \frac{2\omega_0^3}{Q(1+2k)} \\ b_0 &= \frac{\omega_0^4}{1+2k} \end{aligned} \tag{11.17}$$

Assuming $k \ll 1$, the four poles are found to be

$$p_1, p_1^*, p_2, p_2^* = -\frac{\omega_0}{2Q} \pm j\omega_0\left(1 \pm \frac{k}{2}\right) \tag{11.18}$$

To be maximally flat, $k = 1/Q$. Then the gain at band center is $g_mQ/2C\omega_0$ and the bandwidth is $2^{1/2}k\omega_0$. The gain-bandwidth product is thus found to be $g_m/(2^{1/2}C)$ which, when C is half the total interstage capacitance, is $(2)^{1/2}$ times as large as that of the single-tuned circuit.

Assume now that one of the two circuits of Fig. 11.5 has an infinite Q. Let us assume for simplicity that $L_1 = L_2 = L$, $C_1 = C_2 = C$, and either G_1 or $G_2 = 0$. Because either conductance can be set equal to zero with no change in the form of the equations, subscripts on G and Q need not be used. The equations are the same as eqs. 11.16 and 11.17 except that b_1 and b_3 must be halved and the term $\omega_0^2/Q^2(1+2k)$ deleted from the expression for b_2. Then, assuming $k \ll 1$, the poles are

$$p_1, p_1^*, p_2, p_2^* = -\frac{\omega_0}{4Q} \pm j\omega_0\left[1 \pm \frac{[k^2 - (1/2Q)^2]^{1/2}}{2}\right] \tag{11.19}$$

To be maximally flat, it is necessary that $k = 1/(2^{1/2}Q)$. Then, the gain at band center is $2^{1/2}g_mQ/\omega_0C$ and the bandwidth is $2^{3/2}\omega_0/4Q$. The gain-bandwidth product is therefore g_m/C, which can be as much as twice that of the single-tuned circuit, assuming C is half the total interstage capacitance.

When the maximum possible value of Q is desired in any resonant circuit, little or no external loading conductance is added to the circuit. Then, the bulk of the loss resistance occurs in series with rather than in parallel with the inductance (except for some loss in the relatively large plate resistance of the tube). If the Q is at all high, the series resistance can be converted to an equivalent shunt conductance to give a resonant circuit having the same Q and behaving much the same way. (See Sec. 3.6.)

11.3 Note on approximate solutions for pole positions

The method for obtaining the approximate pole positions given in Sec. 11.2 is of sufficient interest to warrant a brief discussion. As an example, we shall consider the equal-Q mutually coupled circuit. We start by assuming that the four poles all have the same real part $-\alpha$ and assume imaginary parts β_1 and β_2. Then, the denominator must factor as

$$(p^2 + 2\alpha p + \alpha^2 + \beta_1^2)(p^2 + 2\alpha p + \alpha^2 + \beta_2^2) \qquad (11.20)$$

The coefficient of p^3 is equal to b_3. Thus

$$4\alpha = b_3 \qquad \alpha = \omega_0/2Q \qquad (11.21)$$

The coefficient of p^0 is equal to b_0 as

$$\alpha^4 + \alpha^2(\beta_1^2 + \beta_2^2) + \beta_1^2\beta_2^2 = b_0 \qquad (11.22)$$

and that of p^2 must be equal to b_2

$$6\alpha^2 + \beta_1^2 + \beta_2^2 = b_2 \qquad (11.23)$$

Substituting between eqs. 11.22 and 11.23

$$\beta_1^4 - \beta_1^2(b_2 - 6\alpha^2) + [b_0 - \alpha^4 - \alpha^2(b_2 - 6\alpha^2)] = 0 \quad (11.24)$$

Solving for β_1^2 and dropping all terms of higher degree in α than the second

$$\beta_1^2 = (b_2/2) - 3\alpha^2 \pm [(b_2/2)^2 - b_0 - 2\alpha^2 b_2 + 4\alpha^4]^{1/2} \quad (11.25)$$

Approximating by neglecting all but the most important terms and assuming $k \ll 1$ in the expressions for b_0 and b_2, and also substituting for α from eqs. 11.21, we get

$$\beta_1^2 = \omega_0^2(1 \pm k) \qquad (11.26)$$

Finally, taking the square root, expanding the radical, and retaining only the most significant term

$$\beta_1, \beta_2 = \omega_0(1 \pm k/2) \qquad (11.27)$$

11.4 Exact factorization for equal-Q circuits

For certain four-terminal interstage networks such as the double-tuned capacitance- or inductance-coupled resonant circuits, an exact factorization of the fourth-degree polynomial in p is possible if the circuits have the same Q and the same resonant frequency.

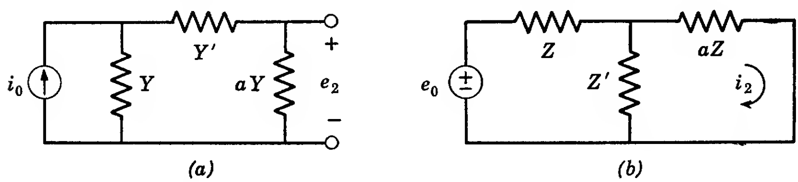


Fig. 11.6. Semisymmetric pi and T networks.

General examples of semisymmetric pi and T networks are shown in Fig. 11.6, for which it should be easy for the reader to verify that

$$\text{Pi:}\quad \frac{e_2}{i_0} = \frac{Y'}{Y[aY + (a+1)Y']} \tag{11.28}$$

$$\text{T:}\quad \frac{i_2}{e_0} = \frac{Z'}{Z[aZ + (a+1)Z']} \tag{11.29}$$

Whenever the two shunt arms of a pi (or the two series arms of a T) network contain the same factor, a partial factorization of the transfer function is possible. In the general pi network, let us set

$$\begin{aligned} Y &= C_1 \frac{p^2 + p(G_1/C_1) + 1/L_1C_1}{p} \\ aY &= C_2 \frac{p^2 + p(G_2/C_2) + 1/L_2C_2}{p} \\ Y' &= pC_3 \quad \text{or} \quad 1/pL_3 \end{aligned} \tag{11.30}$$

In order for Y and aY to have a common factor, we are led to set

$$\begin{aligned} 1/L_1C_1 &= 1/L_2C_2 = \omega_0{}^2 \\ G_1/C_1 &= G_2/C_2 = \omega_0/Q_1 = \omega_0/Q_2 = \omega_0/Q \end{aligned} \tag{11.31}$$

which does not require that C_1 and C_2 be equal. Then

$$\frac{e_2}{e_0} = \frac{-g_m Y_3 p^2}{p^2 + (\omega_0/Q)p + \omega_0^2} \left[\frac{1}{p(C_1 + C_2)Y' + C_1 C_2[p^2 + (\omega_0/Q)p + \omega_0^2]} \right] \tag{11.32}$$

where it has been assumed that $i_0 = -g_m e_0$.

For the inductance-coupled circuit, $Y' = 1/pL_3$; for the capacitance-coupled circuit, $Y' = pC_3$. In either case, the fourth-degree polynomial in the denominator of the transfer function has been broken into the product of two quadratic expressions which can easily be factored. Unfortunately, when one Q is infinite, a general factorization does not appear possible.

11.5 Feedback systems

The number of different arrangements of feedback useful for obtaining band-pass characteristics is almost limitless. We shall discuss a few specific examples, which demonstrate most of the basic types.

The first circuit we shall consider is the feedback pair. This circuit is quite analogous to the low-pass feedback pair discussed in Chap. 10. It consists of two stages of single-tuned networks, each of which has the same center frequency and the same bandwidth. Without feedback, the system becomes a two-stage synchronous-tuned amplifier in which the poles lie at the same points on the p plane. By employing a small amount of resistive plate to plate feedback, these poles are caused to

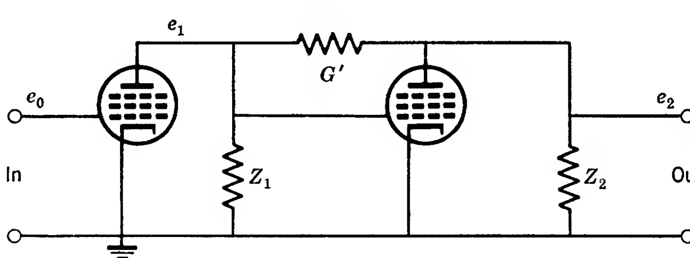


Fig. 11.7. The feedback pair.

separate, giving the effect of two nonidentical single-tuned circuits. The circuit is shown in Fig. 11.7. Let

$$Y_1 = Y_2 = G + pC + 1/pL \tag{11.33}$$

and assume the transconductances of the two tubes to be g_{m1} and g_{m2}

respectively. Then

$$\frac{e_2}{e_0} = \frac{g_{m1}(g_{m2} - G')p^2/C^2}{p^4 + b_3p^3 + b_2p^2 + b_1p + b_0} \tag{11.34}$$

where

$$\begin{aligned} b_3 &= 2(G + G')/C \\ b_2 &= 2/(LC) + (G^2 + 2GG' + g_mG')/C^2 \\ b_1 &= 2(G + G')/(LC^2) \\ b_0 &= 1/(LC)^2 \end{aligned} \tag{11.35}$$

There are two zeros at the origin, as in the two-stage synchronous-tuned amplifier, and four poles. For narrow bandwidths, the poles are at

$$p_1, p_1^*, p_2, p_2^* = -\frac{G + G'}{2C} \pm j\left[\omega_0 \pm \frac{[G'(g_{m2} - G')]^{1/2}}{2C}\right] \tag{11.36}$$

where $\omega_0^2 = 1/LC$. For $G' \ll g_{m2}$ and G, the conductance G' can be neglected in some terms of eq. 11.36. When tuned for maximal flatness, the gain-bandwidth product for the pair is

$$\frac{g_{m1}(g_{m2} - G')}{C^2} \cong \frac{g_{m1}g_{m2}}{C^2} \tag{11.37}$$

which is essentially the same as that of a maximally flat amplifier built up from isolated stages with parallel-resonant circuits. The advantage of the feedback pair is that identical resonant circuits can be used and some negative feedback provided. The principle of the feedback pair can be extended to several stages with feedback applied to all stages as

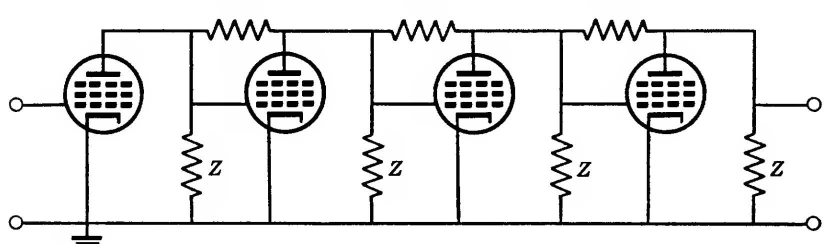


Fig. 11.8. The feedback quadruple.

is indicated in Fig. 11.8. Then, an N-stage amplifier can be made to have a single N-pole maximally flat or similar transfer function.

The second type of feedback band-pass network we shall discuss is a quite different type useful at low frequencies where inductances are not

convenient or may not exist in reasonable form. Consider a circuit consisting of a null network, a tube, and a feedback path. A null network is one furnishing a pair of zeros on or near the $j\omega$ axis. The circuit with a "twin-T" null network is shown in Fig. 11.9. Since our main interest is in low-frequency operation, the distributed capacitances will be neglected. We shall also assume that the plate resistance of the tube

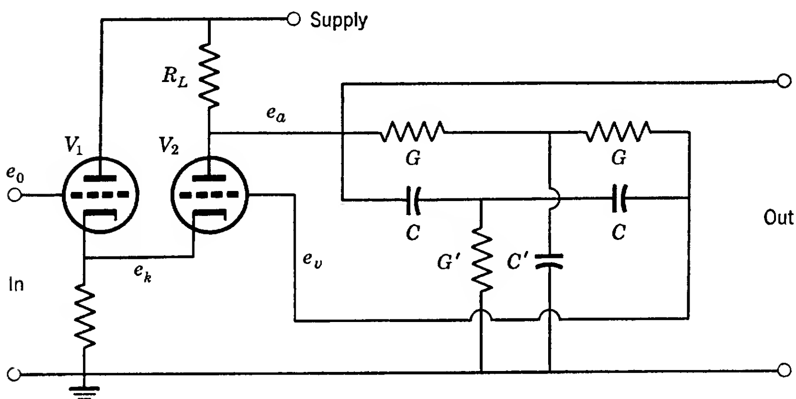


Fig. 11.9. A twin-T feedback amplifier.

is so small and the plate load resistance R_L so large that both can be neglected. Similarly, the interstage coupling network will be neglected. Finally, we shall assume that tube V_1 draws considerably more current than V_2 such that $e_k/e_0 \cong 1$, and that the plate voltage of V_2 is given by

$$e_a = -\mu(e_v - e_k) \cong -\mu(e_v - e_0) \tag{11.38}$$

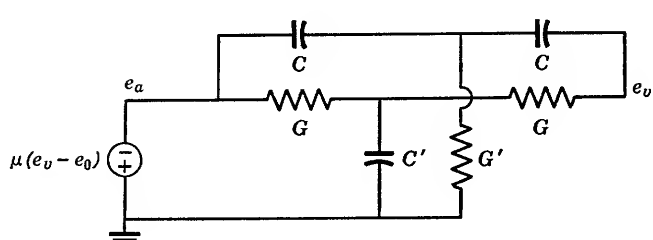


Fig. 11.10. Equivalent circuit of the twin-T feedback amplifier.

With these assumptions, the equivalent circuit takes the form shown in Fig. 11.10. A straightforward application of nodal analysis obtains the characteristics of both the twin-T network and the desired transfer

function e_a/e_0. First, we obtain for the twin-T network alone

$$\frac{e_v}{e_a} = \frac{C^2C'[p^3 + (2G/C')p^2 + (2G^2/CC')p + (G^2G'/C^2C')]}{\text{Denominator}} \tag{11.39}$$

In order to operate as a band-pass amplifier, the twin T must have zeros on the $j\omega$ axis at the resonant frequency ω_0. This causes the system to have no feedback at resonance, whereas at all other frequencies negative feedback reduces the gain. Thus we want the numerator of eq. 11.39 to factor according to

$$(p + a)(p^2 + \omega_0{}^2) = p^3 + ap^2 + \omega_0{}^2p + a\omega_0{}^2 \tag{11.40}$$

By equating the coefficients of the polynomials of eqs. 11.39 and 11.40, we get

$$4G/C' = G'/C \qquad \omega_0{}^2 = 2G^2/CC' \tag{11.41}$$

Both the numerator and denominator of the expression for e_v/e_a will now contain an identical factor which cancels, thereby simplifying the twin-T function. The result is

$$\frac{e_v}{e_a} = F(p) = \frac{p^2 + (2G^2/CC')}{p^2 + [(2G/C) + (4G/C')]p + (2G^2/CC')} \tag{11.42}$$

If negative feedback is to be most effective, the phase shift near the null frequency should average at zero degrees. This means that the

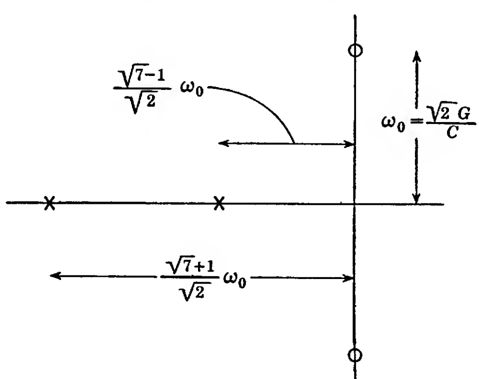


Fig. 11.11. P-z of the twin-T network.

phase shift will be ± 90 degrees just above and just below ω_0. Consequently, we are led to set the real part of the denominator equal to zero

at $p = \pm j\omega_0$. This determines G' and C' uniquely as

$$C' = C \qquad G' = 4G \tag{11.43}$$

The p-z of the resulting null network are shown in Fig. 11.11.

Rather than solve the complete set of equations to find the transfer function e_a/e_0, we can use Fig. 11.10 and eq. 11.42. We have

$$e_a = -\mu(e_v - e_0) \tag{11.44}$$

Substituting for e_v from eq. 11.42, we get

$$\frac{e_a}{e_0} = \frac{\mu}{1 + \mu F(p)} \tag{11.45}$$

We see that the poles of $F(p)$ become the zeros of the transfer function. Substituting for $F(p)$ in eq. 11.45, we get

$$\frac{e_a}{e_0} = \frac{[\mu/(1+\mu)][p^2 + (6G/C)p + (2G^2/C^2)]}{p^2 + \{6G/[C(1+\mu)]\}p + (2G^2/C^2)} \tag{11.46}$$

The poles of e_a/e_0 are at

$$p_1, p_1{}^* = -\frac{3G}{C(1+\mu)} \pm j\frac{(2)^{1/2}G}{C}\left(1 - \frac{9}{2(1+\mu)^2}\right)^{1/2} \tag{11.47}$$

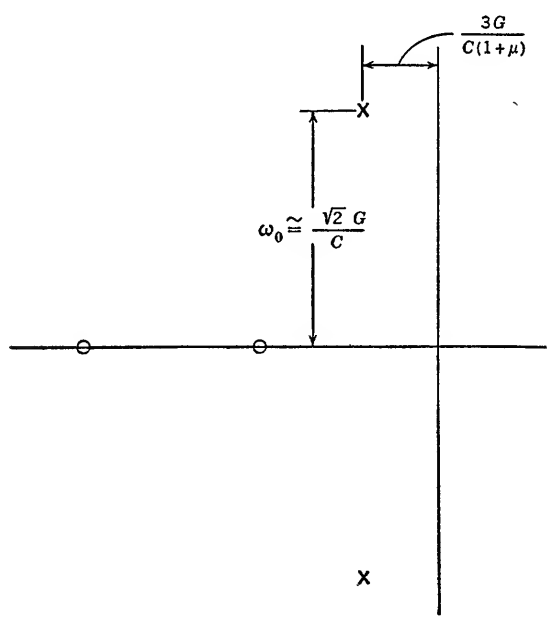


Fig. 11.12. P-z of the twin-T feedback amplifier.

where the second term in the square root is almost always entirely negligible. The p-z of e_a/e_0 are shown in Fig. 11.12, from which the nature of the band-pass behavior of the transfer function is quite evident. At resonance, the gain is seen to be μ. The bandwidth is clearly $(6G/C)/(1 + \mu)$. Large values of μ contribute to high gains and to narrow bandwidths. For a given μ, increasing G/C increases the bandwidth but does not change the gain at resonance nor the resonant frequency.

11.6 Regenerative feedback

Sometimes regenerative feedback can be advantageously employed for obtaining a band-pass characteristic. A rather general equivalent circuit for such systems is shown in Fig. 11.13, where e_v is the applied signal. The output from the network is a voltage e_2, a fraction k of which

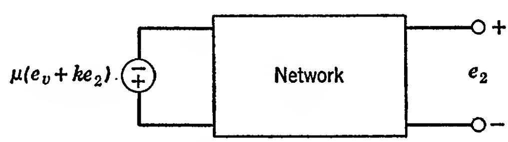


Fig. 11.13. Block diagram of a feedback amplifier.

is added to e_v with the sum applied to the voltage generator. It is assumed that k is real and positive. Actually, Fig. 11.13 is applicable to all kinds of feedback devices including servomechanisms.

For the present example, it will be assumed that the network consists of a two-section R-C filter. The equivalent circuit is shown in Fig. 11.14.

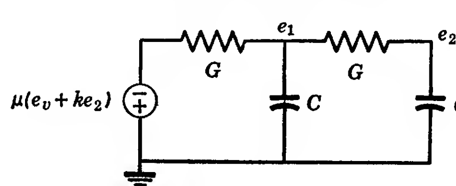


Fig. 11.14. A feedback amplifier employing a low-pass network.

The plate resistance of the tube can be assumed to be part of the conductance G. We find the transfer function assuming e_1 to be the output as

$$\frac{e_1}{e_v} = -\frac{\mu G}{C}\left[\frac{p + (G/C)}{p^2 + (3G/C)p + [G^2(1 + k\mu)/C^2]}\right] \tag{11.48}$$

The p-z of this function are shown in Fig. 11.15.

The amplifier described could also be used as a low-pass amplifier. (If the zero in the transfer function is not desired, the output can be taken as e_2 rather than e_1.) Another *R-C* section could be added to the network. Then the poles could be put very close to the $j\omega$ axis. In fact, they could then be made to lie to the right of the $j\omega$ axis, in which event the circuit would be a "phase-shift" oscillator.

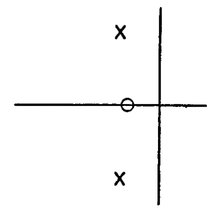


Fig. 11.15. P-z of the feedback amplifier.

Regenerative amplification has a drawback in that the bandwidth may be quite sensitive to changes in the parameters of the tubes (that is, plate resistance, transconductance, and so forth), and for this reason it is not widely favored for stable systems. The circuit of Fig. 11.14 is somewhat useful because it is not excessively sensitive to tube characteristics.

Had a three-section *R-C* network been used, very narrow bandwidths and extremely high gains at resonance could be obtained, but not with any great stability. With three sections and operating the tube in a nonlinear but stable region, the circuit is that of a "regenerative detector."

A regenerative circuit admirably suited for demonstrating different kinds of feedback, as well as one having several practical applications,

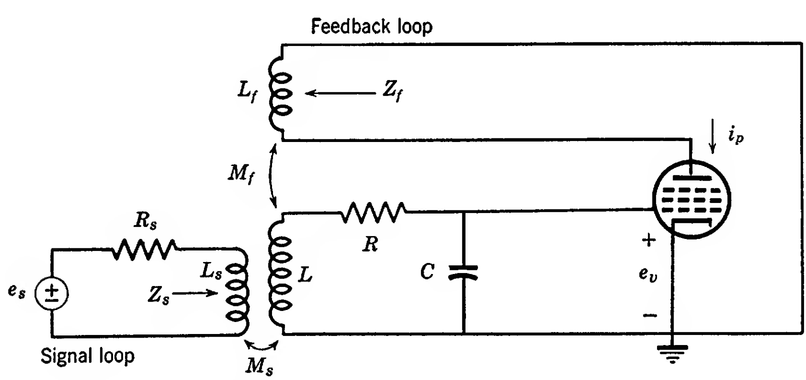


Fig. 11.16. Tuned circuit with feedback.

is shown in Fig. 11.16. The Q of the resonant circuit will be assumed high enough so that no appreciable differences result using either the series resistance or its shunt equivalent.

For purposes of simplifying the ensuing discussion, let us assume that the inductances of the feedback and signal loops are small compared to L so that any capacitances in shunt with them can be neglected. Fur-

ther, assume that the impedance looking into the feedback loop Z_f is much smaller than r_p and also that the impedance looking into the signal loop Z_s is much smaller than R_s. Then the tube and signal sources can both be treated as ideal current generators, and the equivalent circuit becomes as shown in Fig. 11.17.

Because of the assumptions we have made, the voltage induced in L by the signal and feedback circuits may easily be calculated; for the signal it is $\pm G_s M_s p e_s$ and for the feedback it is $\pm g_m M_f p e_v$, where the $\pm$

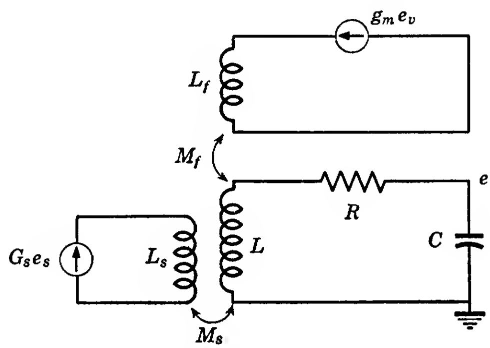


Fig. 11.17. First equivalent of the tuned circuit with feedback.

signs depend upon the connections (or winding directions) of the feedback and signal coils. Finally, we may draw the very simple equivalent circuit shown in Fig. 11.18.

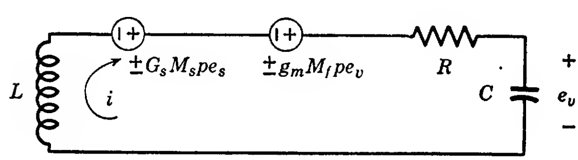


Fig. 11.18. Final equivalent of the tuned circuit with feedback.

In solving for the current i, only one loop equation is required. We get

$$i = \frac{\pm G_s M_s p e_s \pm g_m M_f p e_v}{R + pL + (1/pC)} \tag{11.49}$$

The voltage e_v is given by

$$e_v = (1/pC)i \tag{11.50}$$

Substituting eq. 11.49 in eq. 11.50, we get the transfer function as

$$\frac{e_v}{e_s} = \frac{\pm G_s M_s}{LC}\left[\frac{p}{p^2 + [(RC \mp g_m M_f)/LC]p + (1/LC)}\right] \tag{11.51}$$

which has poles at

$$p_1, p_1{}^* = -\frac{RC \mp g_m M_f}{2LC} \pm j\left[\frac{1}{LC} - \left(\frac{RC \mp g_m M_f}{2LC}\right)^2\right]^{1/2} \tag{11.52}$$

The winding direction of the signal coil merely determines the sign of the transfer function. The winding direction of the feedback coil is

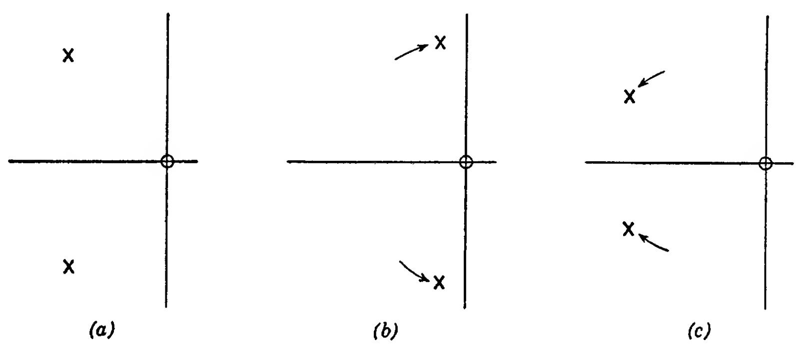


Fig. 11.19. Effect of feedback on the pole locations.

much more important. In fact, for $g_m M_f > RC$, the poles can be made to lie in the right half-plane.

The three feedback conditions are shown pictorially in Fig. 11.19. In *a*, no feedback exists, which occurs when $M_f = 0$. Then, the bandwidth is R/L and the poles lie in the left half-plane. In *b*, there is positive feedback and the poles lie right of their positions with no feedback. The bandwidth, if the system is stable, is

$$B = \frac{R}{L} - \frac{g_m M_f}{LC} \tag{11.53}$$

which is less than that with no feedback. The gain at resonance is

$$\text{Gain} = \frac{G_s M_s}{RC - g_m M_f} \tag{11.54}$$

which is larger than that with no feedback. When the system with positive feedback is stable, the effect of feedback is to increase the Q of

the circuit giving: 1. higher gain at resonance; 2. narrower bandwidth; and 3. higher output impedance (assuming voltage e_v represents the output). When the poles lie quite near the $j\omega$ axis, the gain and bandwidth are quite sensitive to variations in tube and circuit parameters.

If we make M_f large enough so that $g_m M_f = RC$, the gain becomes infinite and the bandwidth zero. Any further increase in M_f will result in oscillations; the circuit is called a "tuned-grid" oscillator and has an oscillation frequency of approximately $(1/LC)^{1/2}$.

If the connections to the feedback coil are reversed so that negative feedback results, quite the opposite effects are obtained as shown by Fig. 11.19*c*. As compared to the system without feedback, we get: 1. lower gain at resonance; 2. wider bandwidth; and 3. lower output impedance. In addition, the transfer function becomes less sensitive to tube characteristics.

When positive feedback is employed but not enough to cause instability, the device is known as a regenerative amplifier and can be made to provide voltage gains of many thousands with quite small bandwidths. If in this condition there exist nonlinearities in the plate-current grid-voltage relationship, detection of a modulated wave can simultaneously be accomplished at the plate; the device is then a regenerative detector.

If the feedback of a regenerative detector is increased to the point where the poles lie very slightly right of the $j\omega$ axis so that the tube oscillates weakly, a carrier input signal near the oscillation frequency will mix with the oscillations to produce an audio-frequency beat note which can be amplified and applied to headphones or a speaker; the regenerative detector can then be used for the reception of keyed continuous-wave signals.

As might be expected, the operation of regenerative amplifiers and detectors is quite critical, the operator of the device often being called upon to make adjustments. For this reason, as well as the sometimes excessively narrow bandwidth, regenerative detectors and amplifiers are not often used in spite of their simplicity and very great sensitivity. A further disadvantage is that moderately strong signals tend to overload them. Actually, the circuit of Fig. 11.16 is only one out of an almost unlimited number that can be used as regenerative amplifiers or detectors.

Suppose the feedback is increased to the point where sustained and strong oscillations can take place. Also suppose the grid of the tube is biased beyond cutoff so that no current flows. Of course, no oscillations can then exist. If the bias voltage is suddenly increased, the tube will build up oscillations exponentially. These oscillations are started by

thermal noise fluctuations and/or switching transients; some variation at the grid must exist in order to start the sequence of feedback events that finally results in sustained oscillations. The stronger the noise, the sooner are these oscillations likely to build up after the plate current is turned on. A signal at the grid will cause the tube to build up oscillations much the same as noise. If the grid is rapidly switched from beyond cutoff to conduction with an external periodic voltage (the "quench" voltage), the waveform at the plate of the tube will consist of repetitive bursts of the sine wave at the oscillation frequency. When a signal is present, each burst builds up sooner; consequently, the pulses of sine wave bursts at the plate will be wider. The waveform at the plate can be passed through a low-pass filter (of bandwidth less than the repetition rate of the quench voltage) to yield the modulation on the input signal. Such a device is known as a separately quenched superregenerative detector. Its bandwidth is about that of the resonant circuit without feedback and its sensitivity is very great. It has the further advantage of discriminating against certain types of interference such as that coming from automobile ignition. An important disadvantage is that the quench frequency must be at least twice that of the highest modulation frequency to be received. Also, since the detector can generate a sine wave of appreciable magnitude, it is often not attached directly to an antenna because of the interfering signal that would be radiated. Rather, it may be decoupled from the antenna with an amplifier stage.

By means of an R-C charging circuit in the grid, the superregenerative detector can be made to generate its own quench voltage; the device is then said to be self-quenched. The sine-wave bursts at the plate for this device are all similar; it is the spacing between them (hence the repetition rate) that is proportional to the magnitude of the applied signal.

One-tube self-quenched superregenerative detectors having high sensitivities (transfer power gains of millions) and being the ultimate in simplicity can be made to operate successfully from frequencies of a few times the quench frequency to far into the microwave region and, in fact, are often used for radar beacons and compact transmitter-receivers.

11.7 Grounded-grid amplifiers

In order to give some indication of methods of analysis applicable to grounded-grid systems, we shall study the cascode band-pass amplifier system in a little detail. The cascode amplifier is important because it has small circuit noise (the smallest of any band-pass amplifier known) and a reasonably high gain, and can be made to operate at relatively

high frequencies. It consists of two tubes which may be either triodes or pentodes, with triodes best for low noise. The amplifier circuit is shown in Fig. 11.20. The first stage has a grounded cathode and the second a grounded grid. The plate load on the first tube is the sum of

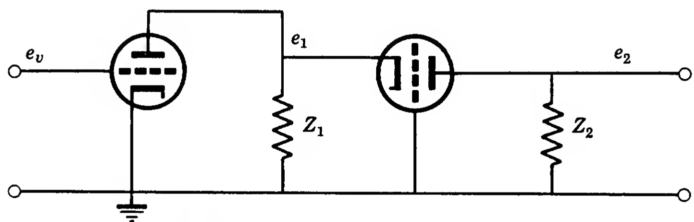


Fig. 11.20. The cascode amplifier.

the circuit load Y_1 and the input admittance of the second tube. With reference to Chap. 9 and assuming that both Z_1 and Z_2 are parallel-resonant circuits, we can draw the equivalent circuit shown in Fig. 11.21, where Y_{in} is the input admittance at the cathode of the second

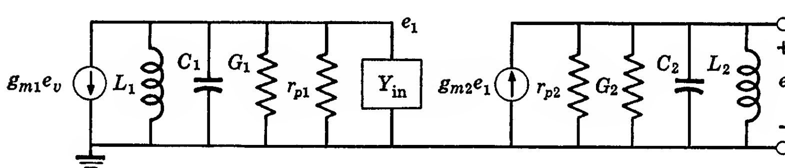


Fig. 11.21. Equivalent circuit of the cascode amplifier.

tube. When the plate circuit of the second tube is resonant

$$Y_{in} = \frac{1 + g_{m2}r_{p2}}{r_{p2} + R_2} \tag{11.55}$$

which is so large that we can consider the load on the first tube to be purely resistive over the most significant parts of the gain function yielded by the tuned circuit load on the second tube. Then, for all frequencies reasonably near the resonant frequency

$$\frac{e_1}{e_v} = \frac{-g_{m1}}{[(1 + g_{m2}r_{p2})/(r_{p2} + R_2)] + G_1 + (1/r_{p1})} \cong -\frac{g_{m1}}{g_{m2}} \tag{11.56}$$

where the approximation is valid for g_{m2} much larger than G_1, G_2, and $1/r_{p2}$, and for $R_2 \ll r_{p2}$.

The ratio of e_2/e_1 is easily found from Fig. 11.21 to be

$$\frac{e_2}{e_1} = \frac{(g_{m2}/C_2)p}{p^2 + \{[G_2 + (1/r_{p2})]/C_2\}p + [1/(L_2C_2)]} \tag{11.57}$$

Combining eqs. 11.56 and 11.57, we get the over-all transfer function as

$$\frac{e_2}{e_v} = -\frac{g_{m1}}{C_2}\left[\frac{p}{p^2 + (G_2'/C_2)p + [1/(L_2C_2)]}\right] \tag{11.58}$$

where $G_2' = G_2 + 1/r_{p2}$. This expression is identical with that of a single grounded-cathode stage with a parallel-resonant interstage network.

Cascode amplifiers can be built with triodes rather than pentodes because the Miller effect is not too pronounced owing to the very low voltage gain (but high power gain) of the first stage (although neutrali-

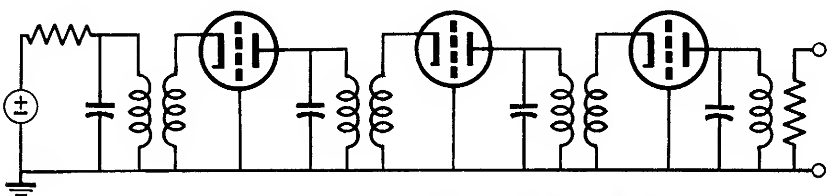


Fig. 11.22. A cascade of grounded-grid stages.

zation may sometimes be required, particularly for low-noise applications). Cathode lead inductance in the second tube is unimportant because of the low impedance level and because it can be made part of the interstage network. In addition, if the grid of the first stage is driven from a low-impedance source, cathode lead inductance in that stage is relatively unimportant. In any event, resonant by-passing of the cathode of the first tube can be achieved without too much trouble because it will be little affected by conditions in the plate circuit.

The cascode amplifier can be constructed with a single tube envelope containing two triodes. Then, the circuit is no more complex or larger physically than a single pentode circuit. In recent years the cascode amplifier has gained fame even to the layman as the input stage of television receivers (following the antenna), although it originally was developed as the input stage of radar intermediate-frequency amplifiers.

When several grounded-grid stages are operated in cascade, desirable results are not achieved unless impedance transformation between the plate of one stage and the grid of the following stage is carried out. This is indicated in Fig. 11.22. Such amplifiers yield moderate power gains from the cathode of one stage to the cathode of the next. Actu-

ally, transformers provide only one means for obtaining impedance transformation. Of course, a transformer equivalent may also be used. In addition, L, T, and pi impedance-matching sections (see Chap. 8) are applicable.

Let us obtain the gain of a grounded-grid stage in a cascade of such stages. We shall assume that all impedances are matched. In Chap. 9 the voltage gain from the cathode to the plate of the stage was found to be $(1 + \mu)/2$. The source is one tube and the load is the following tube. The source impedance is consequently r_p, which is to be matched by means of a transformer to the input impedance at the cathode which, at resonance, is given by eq. 11.55. Thus, the impedance transformation ratio (assuming identical stages) is $r_p[(1 + g_m r_p)/(r_p + R)]$. If the losses in the transformers are small, all loading will be furnished by the input impedance of the tube and the impedance transformation ratio becomes $1/(1 + \mu)$. Thus, the transformer turns ratio should be $1:(1 + \mu)^{1/2}$ and the voltage gain from the plate of one tube to the plate of the next will be $(1 + \mu)^{1/2}/2$. The need for high μ tubes is apparent. Grounded-grid amplifiers (other than in association with the cascode amplifier) are useful as band-pass amplifiers above frequencies of several hundred megacycles per second. Other types are preferred at lower frequencies.

11.8 Some practical matters

When triodes are employed as grounded-cathode amplifiers in a band-pass amplifier and when the voltage gain from grid to plate is at all appreciable, neutralization must be employed in order to keep the system stable. One neutralization method was mentioned in Chap. 9, in which a coil is placed between grid and plate and tuned to resonance with the grid-to-plate capacitance at the center frequency of the amplifier. Unfortunately, this simple method of neutralization is often not adequate as, for example, when it is applied to an amplifier stage that must be tuned. It is not a simple task to tune the neutralizing inductance when the amplifier is tuned because the adjustment is fairly critical. A self-adjusting method is shown in Fig. 11.23, which requires a balanced resonant circuit. The voltages at the two ends of the coil are equal in magnitude and 180 degrees out of phase. Consequently, the positive Miller capacitance from grid to plate can be cancelled with a negative Miller capacitance purposely introduced. This form of neutralization is sensitive to tuning only to a second order and thus need not be adjusted as the amplifier plate circuit is tuned.

Doubling and trebling the input carrier frequency in an amplifier can be accomplished by operating the tube in a nonlinear region of its char-

acteristics. As a doubler, the frequency at the plate will be just twice that at the grid; thus the resonant circuits in grid and plate are tuned to quite different frequencies so that Miller capacitance will not lead to instability and neutralization need not be employed (as long as the bandwidth of the resonant plate load is not too large).

Band-pass amplifiers used with receiving apparatus may often be required to have extremely high voltage gains, perhaps millions. This introduces stability problems. A signal near the output of the amplifier may find its way back to the input by any one of a number of paths and

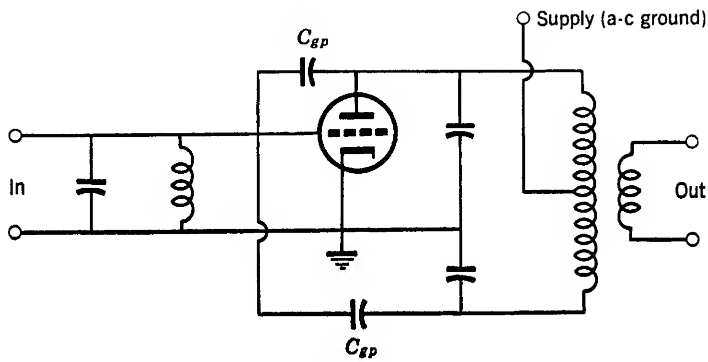


Fig. 11.23. Self-adjusting neutralization.

cause oscillation or misalignment of the tuning characteristic. The troublesome path may be the extremely small mutual inductance or capacitance between input and output. The path may even be furnished by radiation as in an antenna. Such feedback paths can be overcome only with careful shielding and physical separation of input and output. Tubes should have metal shields about them and the stages should be laid out in a straight line. If a voltage gain in excess of a few thousand is needed, a long, narrow, shallow, and completely enclosed chassis should be employed (which acts like a waveguide beyond cutoff). It may be found that a high-impedance input will cause trouble. Reducing the impedance level of the input signal may resolve the difficulty. It is often necessary to thoroughly shield the cables and connectors associated with the amplifier as well. In fact, double-layer shielding of cables is often necessary. If very high gains are desired, it may even be necessary to build two or more amplifiers occupying different chassis with the amplifiers separated by a considerable physical distance. The designer should be well aware of feedback paths whenever he builds an amplifier with a voltage gain above 100 (which can be obtained in some amplifiers with but a single tube).

Undesirable feedback may also be introduced in power-supply wiring, that is, through the screen, plate, filament, and automatic gain control (AGC) leads. Although the source impedance of a power supply is small, it may take only a fraction of an ohm to cause trouble. To combat feedback due to these paths, "graded filters" are often employed with multistage amplifiers having gains in excess of a few hundred. The

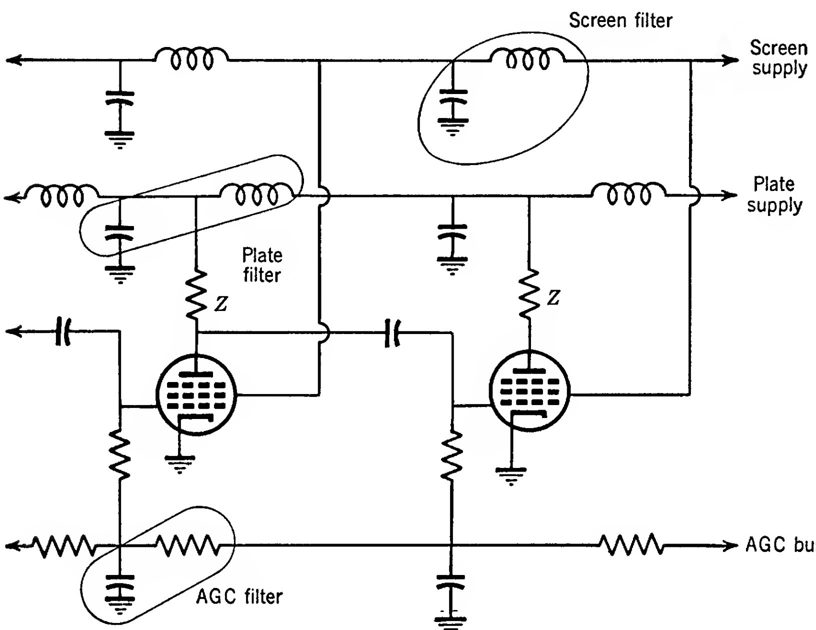


Fig. 11.24. Graded filtering.

power-supply lead is introduced at the output stage and gradually fed to the input stage with an R-C or L-C filter section between each stage. The attenuation through each filter section should be greater than the forward gain of each stage. The general scheme of such filters is shown in Fig. 11.24. The filament leads to the tubes can be filtered in an analogous manner, which may be necessary when the stray capacitance between the filament and other electrodes is not negligible.

Another thing that can cause feedback is the "ground loop." Signal currents can flow in parts of the chassis and therefore flow through a small but nevertheless finite impedance. The resulting induced voltages may introduce trouble. For this reason, the ground connections applicable to each stage should be fastened to the same point so that signal currents in the chassis can be localized as much as possible. A linear

arrangement of tubes and separate ground connections at opposite ends of the chassis for input and output cables helps.

Finally, even though the grid-to-plate capacitance of a pentode is small, it may be large enough to introduce troublesome Miller feedback in a grounded-cathode amplifier, especially if the operating frequency is high. Although the feedback may not be sufficient to result in oscillation, it may be large enough to introduce tuning difficulties by causing the various interstage networks to react upon one another.

Problems

1. Design a synchronous-tuned amplifier having a center frequency of 30 mcs, an over-all bandwidth of 2 mcs, and a voltage gain of 100,000 using tubes with g_m = 5000 micromhos and a gain-bandwidth product of 60 mcs. Show cathode resistors (68 Ω) and suitable by-pass capacitors. Also show suitable plate, screen, and filament graded filters.

2. A typical circuit at the input to a band-pass amplifier is shown in Fig. P.2. Plot the p-z of e_1/e_s and develop suitable design procedures for determining the element values of the circuit given R_s and C. A conjugate match is required to the source resistance R_s.

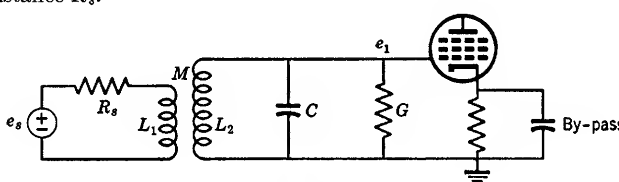


Fig. P.2.

3. Redesign the amplifier of Prob. 1 as a cascade of maximally flat pairs.

4. Redesign the amplifier of Prob. 1 as a cascade of maximally flat triples.

5. The band-pass amplifier in a broadcast receiver makes use of two identical double-tuned equal-Q stages. If g_m = 1500 micromhos, $C_1 = C_2 = 40\ \mu\mu$f, and $L_1 = L_2$, determine L, Q, and k in order that the amplifier have a bandwidth of 10 kcs at 465 kcs where each transformer is identical and maximally flat. Determine the gain at band center. Also, determine the series equivalent of the shunt resistance of the coil.

6. Repeat Prob. 5 assuming one Q of each transformer is infinite.

7. Repeat Prob. 5 using the equal-Q double-tuned capacitance-coupled circuit.

8. Compare the exact and approximate pole locations for the double-tuned mutually coupled and capacitance-coupled circuits. Use equal Q's and equal resonant frequencies.

9. A double-tuned mutually coupled equal-Q circuit with both resonant frequencies equal is used to obtain a pass band of 20 to 35 mcs where $C_1 = C_2 = 20\ \mu\mu$f. The pi equivalent of the transformer is employed. Draw the circuit and label all element values. Use the exact factorization for determining all parameters but use the narrow-band approximation in locating the proper band-pass poles to give a maximally flat function.

10. A capacitance unavoidably exists across the mutual element in the equivalent pi of Prob. 9 resulting in a parallel resonance at 40 mcs. Sketch the magnitude of

the transfer function (first locating the p-z) and compare to that when the bridging capacitance is zero. May the existence of this capacitance be of value in some cases?

11. The transformer-coupled interstage network has a pi equivalent which cannot be realized unless all the inductances are positive. The borderline situation arises when one of the shunt inductances goes to infinity, resulting in the circuit of Fig. P.11. Obtain the transfer function for this network in the form of eqs. 11.5 and 11.6 using $Q_1 = R_1/\omega_1 L_1$, $Q_2 = R_2/\omega_2 L_2$, $\omega_1^2 = 1/L_1C_1$, and $\omega_2^2 = 1/L_2C_2$.

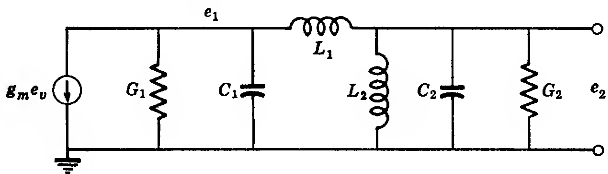


Fig. P.11.

12. Determine the coefficients b_0, b_1, b_2, and b_3 for the maximally flat band-pass function using a double-tuned transformer for a pass band of 20 to 35 mcs.

13. For $Q_1 = \infty$ and $C_1 = C_2 = 20\ \mu\mu f$, determine element values of Fig. P.11 to obtain the transfer function of Prob. 12. Determine the gain at band center and the gain-bandwidth advantage factor.

14. Factor the transfer function of Prob. 11 using the narrow-band approximation.

15. Put the transfer functions for the circuits of Fig. P.15 in factored form.

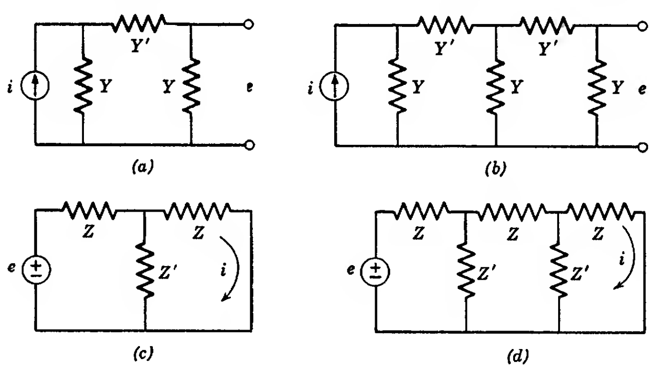


Fig. P.15.

16. Using the results of Prob. 15, plot the locations of the p-z of the transfer function of the circuit of Fig. P.16. Also, determine the equivalent circuit using two identical transformers.

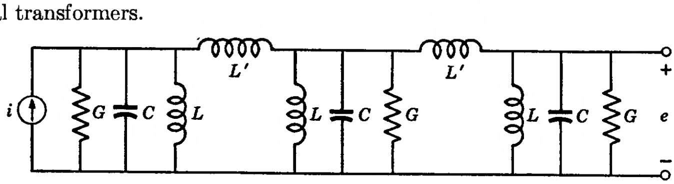


Fig. P.16.

17. Plot the p-z of the transfer function of Fig. P.17.

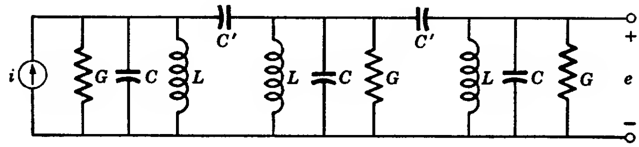


Fig. P.17.

18. Design a feedback pair with a maximally flat pass band of 60 to 65 mcs using tubes with a g_m/C of 40 mcs and $C = 20\ \mu\mu$f.

19. Obtain the design relations for the feedback pair when the load capacitance of the second tube is twice that of the first tube.

20. Determine the design relations for the feedback triple and apply them in the design of the pass band of Prob. 18.

21. Design a twin-T feedback amplifier having as narrow a pass band as is possible at a frequency of 150 cps. Use a tube with $\mu = 70$. Use $R = 500$K in the twin-T which can be assumed to be large compared to r_p and R_L.

22. Use two twin-T feedback amplifiers to get a maximally flat pass band 10 cps wide at 200 cps using $\mu = 70$ and $R = 500$K and the same approximations as in Prob. 21.

23. Design a cascode amplifier using triodes with $g_m = 5000$ micromhos to have a pass band 9 mcs wide centered at 200 mcs. Use $C_1 = C_2 = 18\ \mu\mu$f. (Refer to Fig. 11.20.)

24. Design a four-stage grounded-grid amplifier chain using L-section constant-k matching in the input and for all interstage networks. For the interstage design, neglect the input capacitance of the tube compared to the output capacitance. Use a parallel-resonant circuit at the final output with a bandwidth of 100 mcs. The source resistance is 50 Ω, $\mu = 80$, and $g_m = 12{,}000$ micromhos. Determine the over-all voltage gain and the over-all bandwidth. The center frequency is 400 mcs. Each tube has an output capacitance of 5 $\mu\mu$f and an input capacitance of 15 $\mu\mu$f. Do not add loading to the interstages.

12

Feedback Amplifiers and Stability

We have already treated some feedback devices, notably certain kinds of low-pass and band-pass amplifiers. However, the treatment was limited to simple systems where the transfer function (with feedback) could be factored in general terms so that the positions of the p-z could be determined explicitly. Whenever such is the case, feedback devices offer no special problems, the methods we have already studied being adequate for purposes of analysis and design. One exception to this may occur; the behavior of the system whose p-z are known may admit of improvement by the addition of corrective networks.

We have not discussed devices whose transfer functions are so complex that they cannot be factored in general unless all coefficients are numerical. (Even then, factorization may be impractical.) When such systems must be studied, new methods of attack are practically mandatory. It is the purpose of this chapter to introduce some of these concepts as well as to study certain aspects of stability. It will be found that the p-z concept is a useful adjunct to studies of feedback systems, even though the exact positions of the p-z are not known.

The last topic in this chapter is a discussion of design methods employing p-z techniques. In this, the open-loop transfer function is tailored so that when feedback is applied, the closed-loop behavior follows some *prescribed characteristic*, such as a maximally flat, equal-ripple, or linear-phase function. These procedures are clearly precision design methods. They will be applied again in Chap. 14.

12.1 Generalized single-loop feedback circuits

Most feedback devices can be diagrammed in block form as shown in Fig. 12.1. The input voltage is e_0 and the feedback voltage is e_3. The output may be e_1, e_2, or e_3. For the present, we shall define the charac-

teristics of the boxes of Fig. 12.1 as

$$\begin{aligned}
-K_1G_1 &= -K_1\frac{A_1}{B_1} = -K_1\frac{1 + a_1p + a_2p^2 + \cdots}{1 + b_1p + b_2p^2 + \cdots} \\
K_2G_2 &= K_2\frac{A_2}{B_2} = K_2\frac{1 + a_1'p + a_2'p^2 + \cdots}{1 + b_1'p + b_2'p^2 + \cdots}
\end{aligned} \tag{12.1}$$

which exemplify the special case of finite gain at $\omega = 0$ and imply that the behavior of the system at high frequencies is of most concern (that

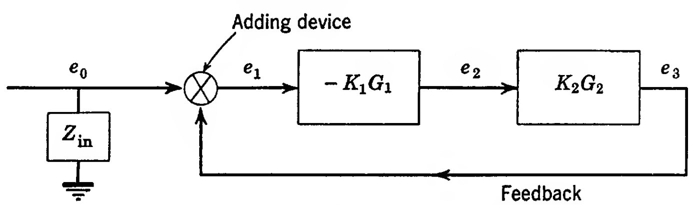


Fig. 12.1. General single-loop feedback circuit.

is, low-frequency p-z, if any, are ignored). The negative signs in eq. 12.1 and Fig. 12.1 permit K_1 and K_2 to be taken as positive numbers, which is a matter of convenience.

Any impedance seen by the input voltage e_0 can be lumped into Z_{in}. If in the physical circuit an impedance appears to be connected between the feedback voltage and ground, its effects can be lumped into the transfer function K_2G_2. As a result, virtually all single-loop feedback devices can be put in the form of Fig. 12.1.

The equations defining the system are

$$\begin{aligned}
e_1 &= e_0 + e_3 \\
e_2 &= -K_1G_1e_1 \\
e_3 &= K_2G_2e_2
\end{aligned} \tag{12.2}$$

Substituting between these equations, we obtain three transfer functions as

$$\begin{aligned}
\frac{e_1}{e_0} &= \frac{1}{1 + K_1G_1K_2G_2} = \frac{B_1B_2}{B_1B_2 + K_1A_1K_2A_2} \\
\frac{e_2}{e_0} &= \frac{-K_1G_1}{1 + K_1G_1K_2G_2} \\
\frac{e_3}{e_0} &= \frac{-K_1G_1K_2G_2}{1 + K_1G_1K_2G_2}
\end{aligned} \tag{12.3}$$

Since the poles of all these functions are the same, stability studies can be made of any one of the three. In general, the transfer function will have zeros as well as poles, and if the degree of B_1B_2 is higher than the second (in all cases it is higher than that of $K_1A_1K_2A_2$), factorization problems will be introduced.

Let us define a composite transfer function as

$$K_1G_1K_2G_2 = KG = K\frac{A}{B} \tag{12.4}$$

where A and B are polynomials with values of unity at $\omega = 0$. Then, eqs. 12.3 become

$$\begin{aligned} \frac{e_1}{e_0} &= \frac{1}{1 + KG} = \frac{B}{B + KA} \\ \frac{e_2}{e_0} &= -K_1G_1\frac{e_1}{e_0} = \frac{1}{K_2G_2}\left(\frac{e_3}{e_0}\right) \\ \frac{e_3}{e_0} &= \frac{-KG}{1 + KG} = \frac{-KA}{B + KA} \end{aligned} \tag{12.5}$$

We shall confine our studies to the transfer functions e_1/e_0 and e_3/e_0 except when K_2G_2 is a simple real number. Then there is no essential difference between the transfer functions e_2/e_0 and e_3/e_0. Should e_2 be the desired output voltage and K_2G_2 frequency sensitive, it can most easily be found by dividing e_3/e_0 by K_2G_2. Usually, K_2G_2 is a fairly simple, passive, and known transfer function.

If the feedback loop of Fig. 12.1 is removed, the transfer function e_3/e_0 becomes

$$\frac{e_3}{e_0} = \frac{e_3}{e_1} = -KG \tag{12.6}$$

which is termed the "open-loop" transfer function, or sometimes the "loop" function. Since this function plays such an important part in eqs. 12.5, it might be expected that its study (that is, the study of the open-loop characteristics) leads to a knowledge of the "closed-loop" system (that is, the system with feedback).

12.2 Open-loop plots and p-z positions

For simplicity let us assume that $K_2G_2 = \beta$, a real number. Then

$$\frac{e_1}{e_0} = \frac{\beta}{1 + K_1G_1\beta}$$
$$\frac{e_2}{e_0} = \frac{-K_1G_1}{1 + K_1G_1\beta} \qquad (12.7)$$

which is typical of many feedback amplifiers. The quantity β is termed the "feedback factor," which represents the fraction of the output voltage fed back and added to the input. The quantity $K_1\beta$ is the amount of feedback and is often expressed in decibels.

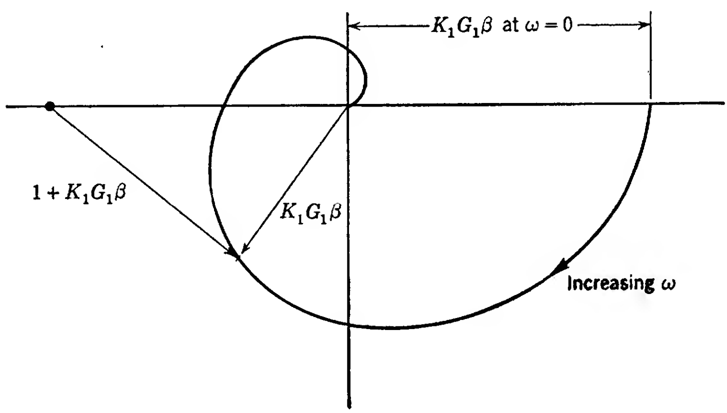


Fig. 12.2. Typical locus plot of a low-pass open-loop system.

The function $K_1G_1\beta$ can be written in terms of the variable $j\omega$ and plotted as a locus on a locus plane. (Note: We shall term $+KG$ rather than $-KG$ the open-loop locus when referring to locus diagrams.) A typical plot is shown in Fig. 12.2 in which the phasor $K_1G_1\beta = KG$ is shown emanating from the origin. Another phasor is shown emanating from the -1 point and terminating on the locus. This second phasor is clearly the quantity $1 + K_1G_1\beta = 1 + KG$, which is the denominator of all the closed-loop transfer functions. The transfer function e_1/e_0 is simply given by the inverse of this phasor. Clearly, the transfer function e_2/e_0 is given by $1/\beta$ (a number) times the ratio of the two phasors shown in Fig. 12.2.

If the open-loop transfer function is stable, then for small β ($\beta = 0$ yielding the open-loop transfer function), the closed-loop transfer func-

tion must also be stable. As β or K_1 is increased, the entire locus is expanded. This causes the phasor $1 + K_1G_1\beta$ to be shorter in length at some frequency. Three cases are exemplified by Fig. 12.3. As $K_1\beta$ is increased, a peak is developed in the value of e_1/e_0 (and e_2/e_0) at the frequency at which the locus comes closest to the -1 point. This means that the closed-loop transfer function has a pole to the left of the $j\omega$ axis but near it. If the value of $K_1\beta$ is increased to a critical value, the locus will pass *through* the -1 point. Then, the phasor $1 + K_1G_1\beta$ becomes zero in length at some frequency and the closed-loop function e_1/e_0 has a pole exactly on the $j\omega$ axis. Any further increase in β will cause the locus

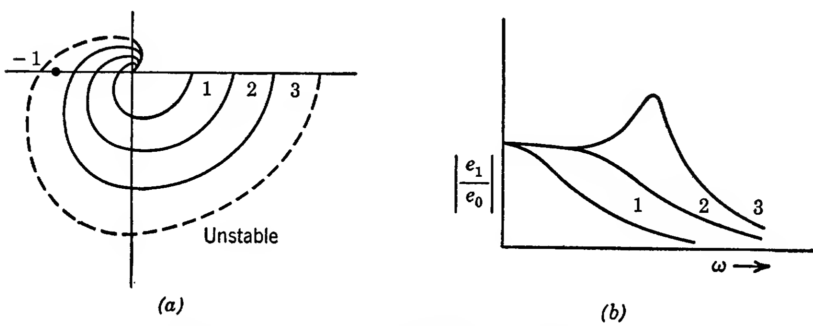


Fig. 12.3. Effects of changing the feedback $K\beta$.

to *encircle* the -1 point, which is equivalent to the pole being moved to the right half-plane to give an unstable system, that is, an oscillator. From this it can be seen that for typical open-loop functions the critical frequency is that where the phase shift of the open-loop transfer function is 180 degrees. The system with feedback will oscillate if the open-loop gain K_1G_1 times the feedback factor β is equal to or larger than unity at this frequency. If it is desired that the system be stable, it is necessary that the loop gain at the 180-degree phase-shift frequency be less than unity.

The open-loop value of e_2/e_0 at $\omega = 0$ of the system of Fig. 12.1 and eqs. 12.7 is $-K_1$ (where G_1 and G_2 are both defined as unity at $\omega = 0$). The gain at $\omega = 0$ with feedback is $1/\beta$ for $K_1\beta \gg 1$; the gain of the system is then fairly independent of the open-loop gain K_1 and consequently is fairly insensitive to the variations of vacuum tubes. This is one of the advantages of negative feedback. If all the benefits of feedback are desired, it is necessary that $K_1\beta$ be quite large. If $K_1G_1\beta$ is to be maintained less than unity at the 180-degree phase-shift frequency while $K_1\beta$ is to be made much larger than unity, rather severe restrictions are imposed upon the open-loop transfer function. This will be considered in more detail later.

The transfer function e_2/e_0 from eqs. 12.7 is often written in a slightly different form, which helps simplify numerical and graphical manipulations. By dividing the numerator and denominator of eqs. 12.7 by $K_1\beta$, we get

$$\beta \frac{e_2}{e_0} = \frac{-G_1}{(1/K_1\beta) + G_1} \tag{12.8}$$

The phasors G_1 and $(1/K_1\beta) + G_1$ are shown in Fig. 12.4. To work from this plot, the locus of G_1, which is normalized to unity gain at

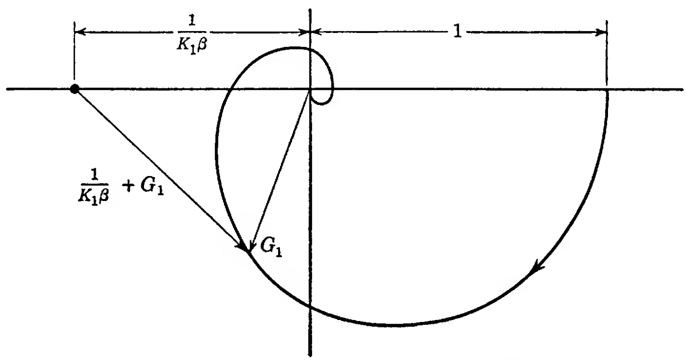


Fig. 12.4. Locus plot normalized to gain.

$\omega = 0$, need be drawn only once. All changes in K_1 or β can be handled by merely moving the point located at $-1/K_1\beta$.

As a more specific example of what we have been talking about, recall the null-network feedback amplifier described in Chap. 11 which employs a twin-T circuit. It can be diagrammed as in Fig. 12.5, where

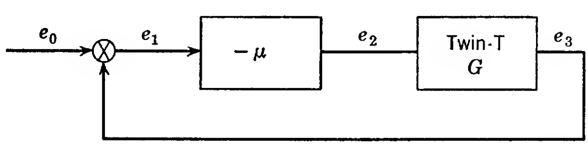


Fig. 12.5. Block diagram of the twin-T feedback amplifier.

e_2 represents the output and μ is the amplification factor of the tube. Except for a constant multiplier μ, we may consider e_1 to be the output.

Various locus plots for G are shown in Fig. 12.6. In Fig. 12.6*a* the zero lies on the $j\omega$ axis and the phase shift is ± 90 degrees at frequencies near the zero. In *b* the zero is also on the $j\omega$ axis but the phase shift is not ± 90 degrees. In this case it can be observed that the transfer func-

tion e_1/e_0 as given by the reciprocal of the phasor shown in Fig. 12.6*b* is not symmetric about the center frequency. In *c* the zero lies slightly left of the $j\omega$ axis. Then, although the phase shift is correct, the bandwidth with feedback becomes larger and the gain at the center frequency smaller as compared to that when the zero lies on the $j\omega$ axis. In *d* the zero lies slightly right of the $j\omega$ axis (the function G is no longer minimum phase), which causes the function with feedback to have a higher

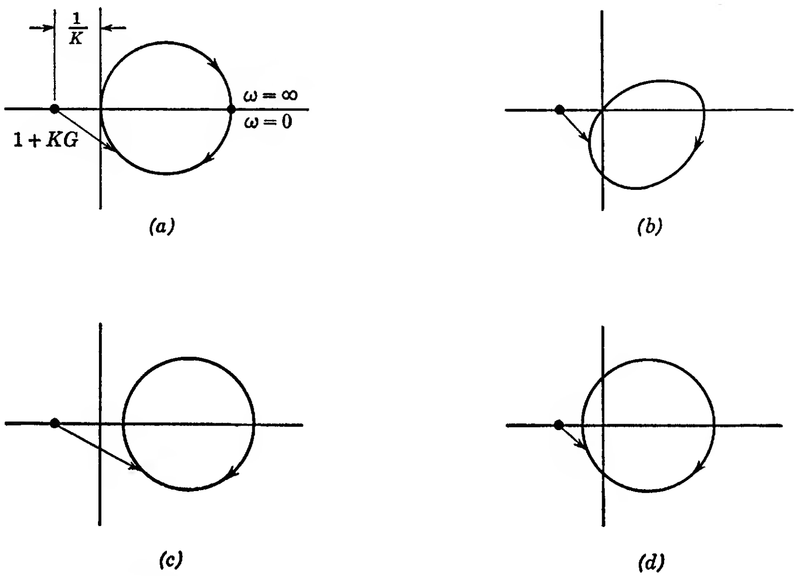


Fig. 12.6. Locus plots for the twin-T feedback amplifier.

gain and a smaller bandwidth as compared to that when the zero lies exactly on the $j\omega$ axis. In this case the system has some positive feedback at the center frequency, whereas when the zero lies on the $j\omega$ axis there is no feedback at the center frequency.

12.3 The general effects of negative feedback

Consider the feedback circuit of Fig. 12.7 in which a disturbing voltage e_x is added at a point within the system. It will be assumed that there exists voltage gain both prior to and after the point where e_x is added. Voltages e_1 and e_3 are given by

$$\begin{aligned} e_3 &= -K_1G_1K_2G_2e_1 + K_2G_2e_x \\ e_1 &= e_0 + \beta e_3 \end{aligned} \tag{12.9}$$

Combining these expressions

$$e_3 = \frac{-K_1G_1K_2G_2}{1 + K_1G_1K_2G_2\beta}\, e_0 + \frac{K_2G_2}{1 + K_1G_1K_2G_2\beta}\, e_x \tag{12.10}$$

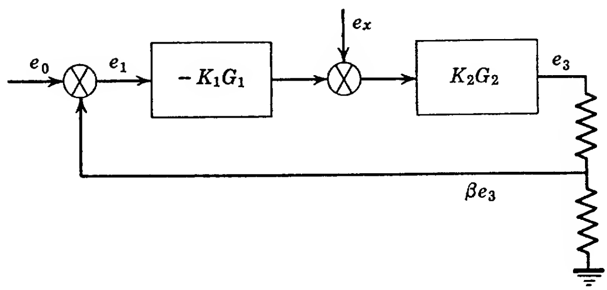


Fig. 12.7. Feedback system with internal disturbance.

At moderately low frequencies, $G_1 \cong G_2 \cong 1$ and eq. 12.10 is approximately

$$e_3 \cong \frac{-1}{\beta + 1/K_1K_2}\, e_0 + \frac{1}{K_1\beta + 1/K_2}\, e_x \tag{12.11}$$

which, for K_1 and K_2 large and β not too small, is approximately

$$e_3 \cong -\frac{e_0}{\beta} + \frac{e_x}{K_1\beta} \tag{12.12}$$

From this it can be seen that the effect of the disturbing voltage e_x at the output is reduced by the factor $1/K_1\beta$, and the gain with feedback is dependent upon the feedback factor β only. If K_1, the gain *prior to* the point where the disturbance is introduced, is large, the effect of e_x at the output may be entirely negligible. Thus we conclude that hum or other noise added to a feedback amplifier after the input voltage has undergone some amplification will be greatly reduced by feedback. By the same token, disturbing voltages introduced prior to the gain $-K_1$ will not be reduced at the output. In other words, *feedback has no benefits when disturbances occur at the input.* As a result, the input stage or stages of an amplifier to which feedback is to be applied are often rather carefully built. The designer may even go so far as to use precision wire-wound resistors and other high-quality components in the early stages of feedback amplifiers designed to transmit direct current because instability in tube operating points and circuit elements is equivalent to low-frequency disturbances.

Also roughly equivalent to a disturbing influence e_x are nonlinearities introduced by vacuum tubes. Since nonlinearities are generally associated with the output stage or stages of an amplifier where signal levels are large, negative feedback can do much to reduce them. Of course, feedback has no beneficial effects when the nonlinearities are excessive, as when a tube saturates.

Consider now the feedback device shown in Fig. 12.8. Here we have shown an equivalent source and source resistance R_s within the feed-

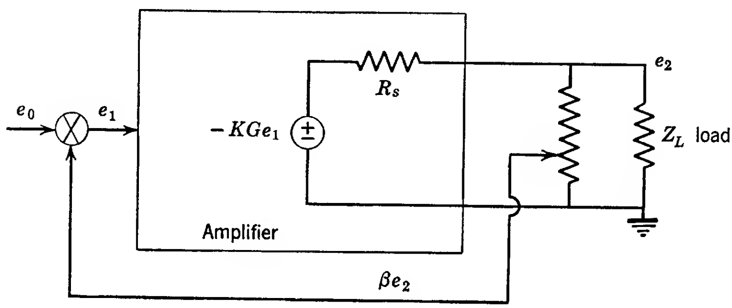


Fig. 12.8. Feedback system with output impedance.

back loop. Z_L is the load applied to the system. It is our purpose to determine the effective output impedance of the amplifier with feedback. Without feedback, the output impedance is obviously R_s. The transfer function is found to be

$$\frac{e_2}{e_0} = \frac{-[(KGZ_L)/(R_s + Z_L)]}{1 + [(KG\beta Z_L)/(R_s + Z_L)]} \tag{12.13}$$

For $Z_L = \infty$, the transfer function is

$$\left(\frac{e_2}{e_0}\right)_{Z_L=\infty} = \frac{-KG}{1 + KG\beta} \tag{12.14}$$

If Z_L is made equal to the equivalent output impedance of the amplifier with feedback Z_{out}, the value of e_2/e_0 will be precisely half that at $Z_L = \infty$. Thus

$$\frac{-KG}{2(1 + KG\beta)} = \frac{(-KGZ_{\text{out}})/(R_s + Z_{\text{out}})}{1 + [(KG\beta Z_{\text{out}})/(R_s + Z_{\text{out}})]} \tag{12.15}$$

Solving for Z_{out}, we get

$$Z_{\text{out}} = \frac{R_s}{1 + KG\beta} \tag{12.16}$$

from which we see that the output impedance with feedback may be considerably less than that without feedback. At low frequencies where $G \cong 1$, $Z_{\text{out}} \cong R_s/(R_s + K\beta)$, which may be a very small resistance. However, at high frequencies, $KG\beta$ must become small in the interests of stability. At high frequencies, feedback therefore has little effect in reducing the output impedance. Likewise, it has little effect in reducing the severity of the various kinds of disturbances and nonlinearities within the amplifier if these disturbances represent high-frequency phenomena.

It should be observed that when the load impedance is known it can be included as part of the circuitry inside the feedback loop and its presence accounted for without necessitating a separate analysis; the transfer function KG is replaced with a new function $KGZ_L/(R_s + Z_L) = K'G'$, where K' is the value of this new function at $\omega = 0$, and G' is unity at $\omega = 0$. If Z_L has a shunt resistance in part, K' will be less than K. Clearly, the effect of load Z_L is small if the open-loop output impedance R_s is small. If an amplifier must operate under a variety of load conditions, it is wise to make the output stage of the amplifier have a low output impedance, as when the output stage in itself is some type of feedback amplifier, such as a cathode follower or a grounded-cathode tube with resistive plate-to-grid feedback. The function G' has a frequency sensitivity that is dependent upon the reactance of the load. This can have rather serious consequences if R_s is not small. It is not unusual for an amplifier that is stable with Z_L purely resistive to become unstable with a capacitive load.

So far, we have talked about voltage feedback in which a sample of the output voltage is used as a feedback quantity. We may also use current feedback in which the feedback voltage is proportional to the load current. Feedback then acts to stabilize the load current rather than the load voltage. The same beneficial effects in regard to reducing sensitivity to noise and other disturbances as well as to nonlinearities are obtained with current feedback. However, whereas voltage feedback causes the output of a feedback amplifier to behave more like an ideal voltage source (as a cathode follower), current feedback tends to cause the amplifier to behave more like an ideal current source (as a grounded-cathode pentode). Consequently, the equivalent output impedance of an amplifier with current feedback may be extremely large.

12.4 High-frequency stability

Consider a typical feedback device having finite gain at $\omega = 0$ and a real feedback factor β as shown in Fig. 12.9. The transfer functions of interest are

$$\frac{e_1}{e_0} = \frac{1}{1 + KG\beta}$$
$$\frac{e_2}{e_0} = \frac{-KG}{1 + KG\beta} \tag{12.17}$$

Suppose a rather conventional amplifier is used to provide the open-loop function. Then all the p-z of $-KG\beta$ (due to stray capacitances,

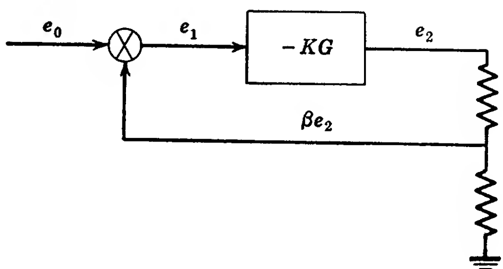


Fig. 12.9. A simple feedback system.

and so forth) will lie roughly the same distance from the origin of the p plane as indicated in Fig. 12.10a to give a locus plot as shown in Fig. 12.10b. If the number of poles in excess of zeros is at all large (three or

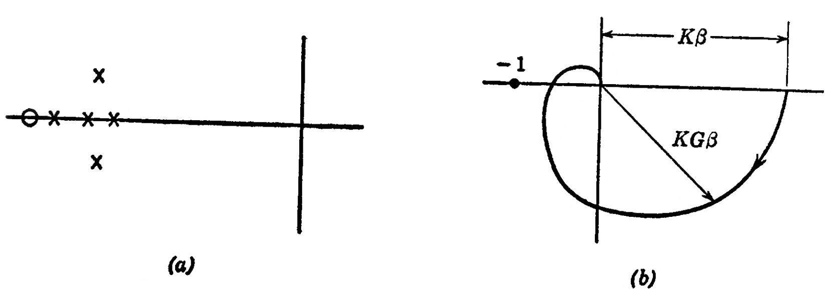


Fig. 12.10. Typical uncompensated amplifier characteristics.

more), it will be found that very little feedback (small $K\beta$) is permissible without introducing an excessive high-frequency peak or complete instability. Since in the limit the poles of Fig. 12.10a may be due solely

to stray capacitance (and perhaps the leakage inductance of a transformer), little can be done to increase the amount of feedback *without purposely narrowing the open-loop bandwidth.* From this we conclude that the benefits of feedback can never be obtained over a bandwidth larger than that realizable without feedback.

The simplest and most foolproof way of designing a feedback amplifier which is stable with large amounts of feedback is simply to shunt some signal point in the system to ground with a single relatively large capacitor so that the ratio of the open-loop bandwidths before and after intentional "bandwidth degradation" of the open-loop function is large. Then, the open-loop p-z plots and loci are as shown in Fig. 12.11. Simply, the modified open-loop system acts like a one-pole function (which can never be unstable) for all frequencies below that where the phase shift contributed by the high-frequency poles becomes appreciable. The result, of course, is to greatly decrease the gain where the phase shift is 180 degrees, which allows $K\beta$ to be increased correspondingly. The open-loop bandwidth of the degraded system should be about the same as that required by the signals to be amplified with the feedback amplifier if the maximum benefits of feedback are desired. Very approximately, if a 100-kcs bandwidth is desired over which feedback is fully effective and if it is desired that $K\beta = 100$, the bandwidth of the system before degradation should be at least $100 \times 0.1 = 10$ mcs.

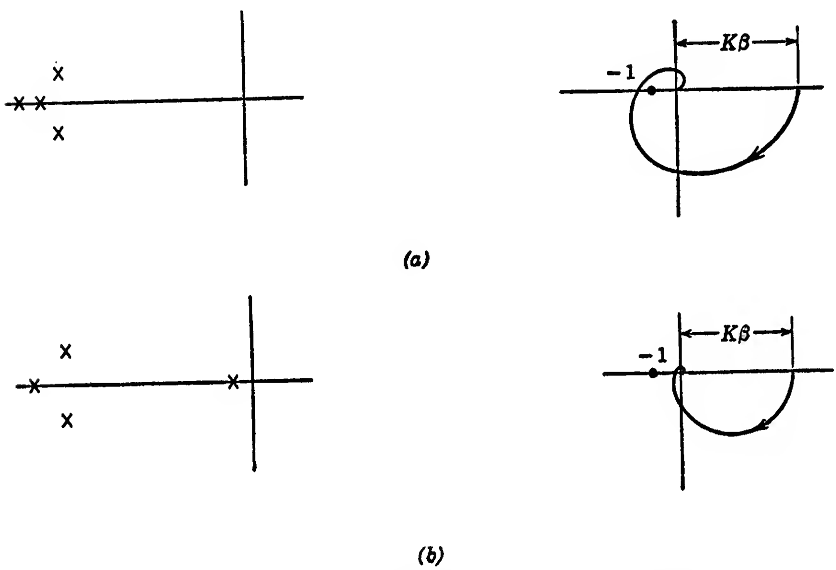


Fig. 12.11. Use of a single pole for bandwidth degradation. (*a*) Before. (*b*) After.

The problems associated with obtaining both large bandwidth with feedback and large $K\beta$ should be noted. Of course, the actual bandwidth with feedback will be much larger than the bandwidth of the degraded open-loop system; however, feedback is not fully effective over the entire bandwidth with feedback because KG decreases with frequency rapidly above the cutoff frequency of the degraded stage.

Any one of many two-terminal interstage networks is an alternative to the single capacitor for degrading the bandwidth in order to achieve

Fig. 12.12. Use of two poles for bandwidth degradation.

the same general effects as given with the capacitor. The reason such networks can be used is that the phase shift is always bounded by 90 degrees. The shunt-peaked interstage network ($\delta \leq 1.554$) is one well suited to this application.

A slightly more complex method for increasing $K\beta$ is to decrease the bandwidth at two points in the circuit. Then the p-z plot and locus are as shown in Fig. 12.12. With this method, an increase in the open-loop bandwidth for the same $K\beta$ as compared to the previous method can sometimes be obtained. If the phase shift contributed by the high-frequency poles is appreciable, this method may not be as good as that previously described (unless zeros are introduced, as will be described later).

Figure 12.13 shows how circuits such as the shunt-peaked one act to nullify the effects of high-frequency poles on the negative real axis of the p plane by moving the poles up the imaginary axis. Until the frequency is so high that it is nearly equal to the distances of the complex poles from the origin, the phase shift contributed by the high-frequency poles is quite small; by the time their phase contribution is appreciable, the poles near the origin have reduced the gain considerably.

A quite common stabilizing procedure in feedback amplifiers is that of adding some capacitance from output to input, which gives the feedback voltage a leading component and is somewhat equivalent to putting a pole of the open-loop function at low frequencies. It has the

advantage of lessening saturation problems. A very small capacitor goes a long way in this regard.

Frequently, the designer of a feedback amplifier is forced to accept a certain type of behavior to begin with. Most commonly, he must deal with a circuit that may be narrower than he desires. For example, it

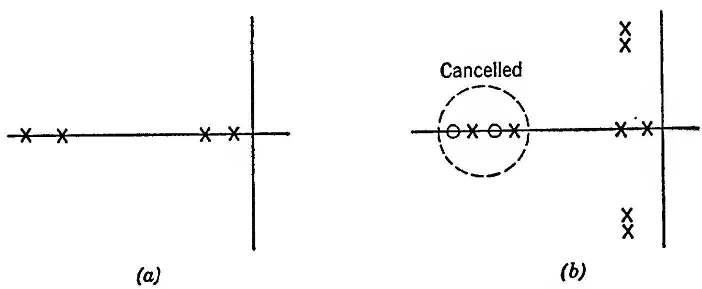


Fig. 12.13. Phase-shift reduction with the shunt-peaked function.

may be required to drive a load containing a large shunt capacitance, or a transformer with a large leakage reactance, or perhaps an electromechanical device which is relatively ponderous. If the designer is fortunate, the low-frequency poles may already be in the optimum positions, thereby allowing him to use the maximum amount of feedback without corrective networks. As is more often the case, the unavoidable pole is too near the origin. When this is so, some type of lead network in cascade with the amplifying system is suggested, as exemplified by Fig. 12.14.

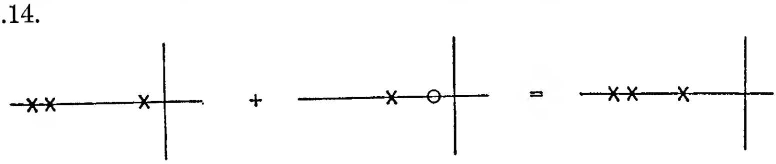
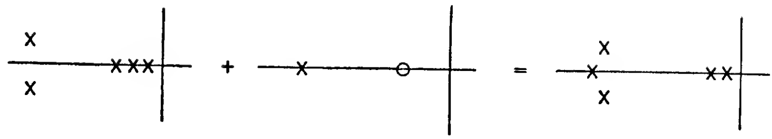


Fig. 12.14. Increasing bandwidth with a lead network.

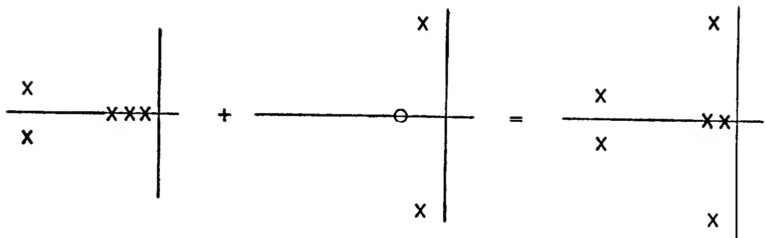
Another situation where a lead network is useful is where too many low-frequency poles exist. Two different approaches to this problem, as given by different kinds of lead networks, are indicated in Fig. 12.15.

Sometimes some moderate-frequency pole is not near enough to the origin. Then a lag network is called for as shown in Fig. 12.16.

Frequently, it is desirable to have a flat gain characteristic with feedback. Little we have said so far indicates methods for achieving this. If the transfer function with feedback is not too complex, it can be equated to a desired transfer function and the coefficients of the powers



(a)



(b)

Fig. 12.15. Use of lead networks for pole removal.

of p equated in order to determine the necessary design relations. However, this procedure becomes unmanageable if the number of p-z is large.

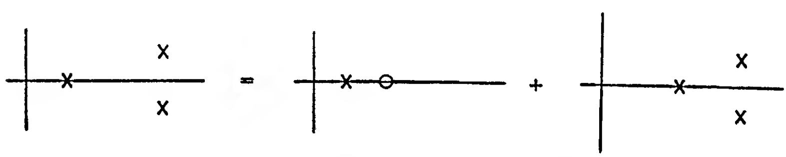


Fig. 12.16. Decreasing bandwidth with a lag network.

The general transfer function for the single-loop system with purely real β can be multiplied by β to give

$$\beta \frac{e_2}{e_0} = \frac{KG\beta}{1 + KG\beta} \tag{12.18}$$

It is possible to find contours on the locus plane which are curves along which the magnitude of eq. 12.18 is constant. If the open-loop function $KG\beta$ can be made to follow one of these curves, the gain with feedback will be constant with frequency. The technique we are about to discuss is valid only if the feedback factor is real.

Let $KG\beta$ be expressed in terms of its real and imaginary parts as

$$KG\beta = R + jX \tag{12.19}$$

We want

$$\left| \frac{KG\beta}{1 + KG\beta} \right| = \left| \frac{R + jX}{1 + R + jX} \right| = \text{Constant} = (k)^{1/2} \tag{12.20}$$

or

$$\frac{R^2 + X^2}{(1 + R)^2 + X^2} = k \tag{12.21}$$

from which

$$X^2 + \left(R - \frac{k}{1 - k}\right)^2 = \frac{k}{(1 - k)^2} \tag{12.22}$$

Equation 12.22 describes circles whose centers are on the real axis at $R_0 = k/(1 - k)$ and which have radii $(k)^{1/2}/(1 - k)$. The constant k

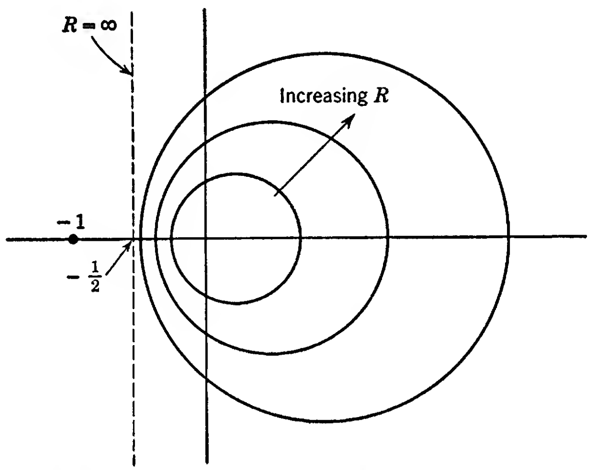


Fig. 12.17. Circles of constant gain with feedback.

will be equal to or less than unity. Some of these circles are indicated in Fig. 12.17. The two intercepts of the circles with the real axis are

$$R_1, R_2 = R_0[1 \pm (1 + 1/R_0)^{1/2}] \tag{12.23}$$

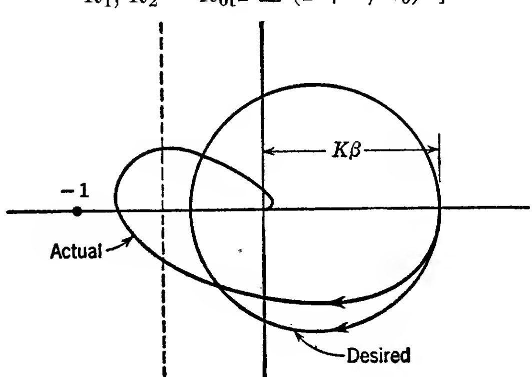


Fig. 12.18. Actual and ideal loci.

To use the constant-gain loci, the intercept farthest out on the real axis is equated to the desired value of $KG\beta = K\beta$ at $\omega = 0$. The resulting situation is depicted in Fig. 12.18, in which both the desired locus (for constant gain with feedback) and the actual system locus are shown.

12.5 Low-frequency stability

Many feedback amplifiers need not transmit very low frequencies; consequently, interstage coupling circuits and interstage and output transformers can be employed. Such circuits (as well as by-pass circuits) introduce low-frequency p-z with the result that excessive leading phase shift may occur at low frequencies. For example, four stages

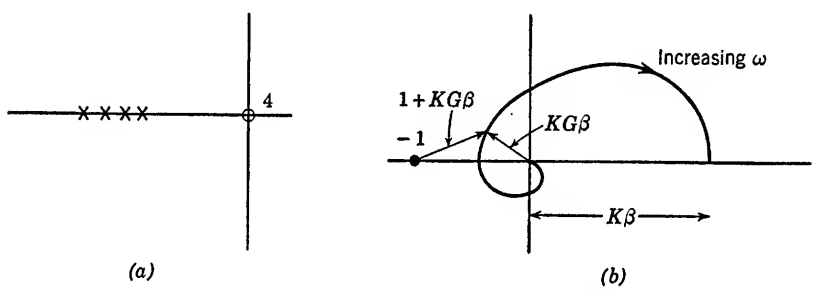


Fig. 12.19. Behavior of four coupling networks.

of R-C coupling networks give a locus somewhat like that shown in Fig. 12.19. It is assumed that the gain at high frequencies (neglecting high-frequency p-z) is $K\beta$. Clearly, for small $K\beta$, the system with feedback must be stable if the open-loop function is stable. However, for

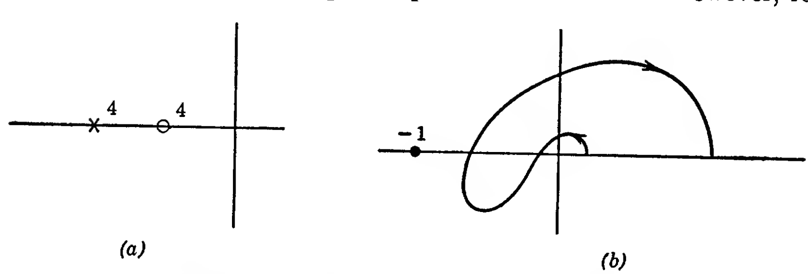


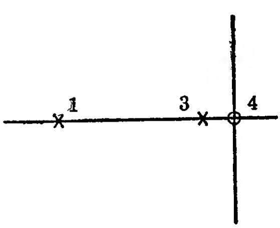
Fig. 12.20. Behavior of four by-pass networks.

larger $K\beta$, the transfer function can develop a peak at low frequencies corresponding to a pole in the left half-plane near the $j\omega$ axis. Causing $K\beta$ to increase even further results in the pole moving across the $j\omega$ axis. Then the function of Fig. 12.19 is that of a type of phase-shift oscillator.

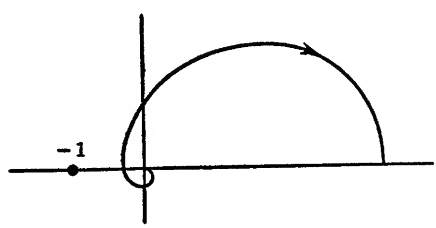
Cathode and screen by-pass circuits also introduce a leading phase shift at low frequencies as, for example, several similar by-pass circuits shown in Fig. 12.20.

An amplifier consisting of several stages can have a large number of low-frequency p-z, perhaps as many per stage as are contributed by two by-pass networks and one coupling network. With *R-C* plate loads, each stage will contribute only one high-frequency pole. Consequently, the problem of low-frequency stabilization can be considerably more difficult than that of high-frequency stabilization. The problem is heightened by the difficulty of obtaining reasonably simple low-frequency compensating networks that furnish complex p-z. The result of all this is that screen and cathode by-pass networks in a feedback amplifier are often either not used or are cancelled with lag networks (for example, low-frequency compensating circuits and cathode peaking to avoid cathode by-passing without loss in voltage gain).

To solve the problem of the zeros furnished by coupling networks and perhaps the magnetizing inductance of a transformer, a technique analogous to the single capacitance shunt for obtaining high-frequency stability can be employed. That is, all the coupling networks except perhaps one or two are made to have little effect down to very low frequencies. The low-frequency cutoff point of one or two of the coupling networks is made as large as is permitted by the design specifications of the feedback amplifier. The result is a low-frequency p-z collection something like that shown in Fig. 12.21. The coupling network having the pole



(*a*)



(*b*)

Fig. 12.21. Stabilization of coupling networks.

most distant from the origin forces the open-loop gain to a low value before the total leading phase shift becomes 180 degrees.

Much of the problem of low-frequency stability can be avoided by making part of the amplifier direct coupled, as is fairly common practice (although certainly not economical unless absolutely necessary). Of course, in feedback amplifiers designed to transmit direct currents, the problem does not arise.

The technique used for improving the long-time response of video amplifiers is also applicable in feedback amplifiers. That is, the sum of the low-frequency pole positions of the closed-loop transfer function can be made equal to the sum of the low-frequency zero positions. However, this does not also guarantee stability, which means that considerations of degrading as discussed before must also be kept in mind.

The most economical way to obtain low-frequency compensation is to use a complex feedback β. If a relatively large capacitor is placed in series with the feedback path (and if the feedback path also has resistive loading), then β becomes $\beta_0 p/(p + a)$, where β_0 is real and is the value of this complex feedback at all but very low frequencies. For a given value of a, the system is made stable at low frequencies by suitable degradation. Then the parameter a can be adjusted to give optimum low-frequency compensation. This type of complex feedback amounts to an effective low-frequency compensating circuit. The design details are left to the reader as an exercise.

12.6 Linear range and intermodulation

All practical amplifying devices have definite maximum voltages or currents that can be obtained. It is not possible to do more than cut off a tube or drive its grid more than slightly positive. A stage of an amplifier (or several stages) without feedback will have a curve of input versus output that has a shape something like that shown in Fig. 12.22

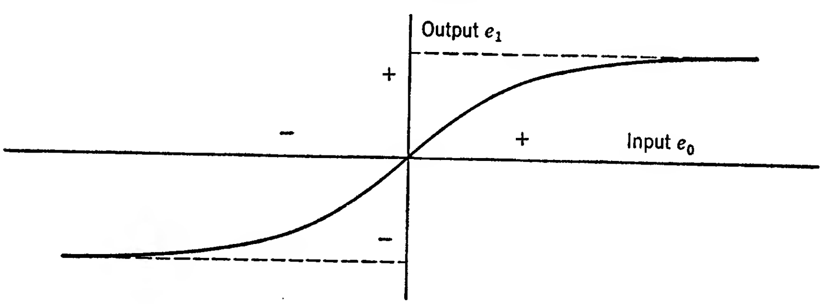


Fig. 12.22. Typical input-output characteristics.

(neglecting frequency-varying effects). No matter how large the input is and how much feedback is applied, the output is definitely bounded at the limits of saturation.

The output-input relationship can be expanded in a Taylor series about zero a-c signal level to give

$$e_1 = \frac{de_1}{de_0} e_0 + \frac{1}{2!}\left(\frac{d^2e_1}{de_0{}^2}\right) e_0{}^2 + \frac{1}{3!}\left(\frac{d^3e_1}{de_0{}^3}\right) e_0{}^3 + \cdots \tag{12.24}$$

The first term of this series represents the linear output-input relationship. The first derivative is the gain, which is the slope of the input-output characteristic at $e_0 = 0$, the second derivative gives the curvature of the input-output characteristic. All the higher derivatives (second and above) measure the amount that the input-output characteristic deviates from linearity. Because of saturation limits, all of these higher order derivatives can never be zero.

If e_0 is a signal represented with a collection of sine waves as a Fourier series or integral, the components e_0^2, e_0^3, and so forth, in eq. 12.24, when expanded, will show additional harmonically related frequency com-

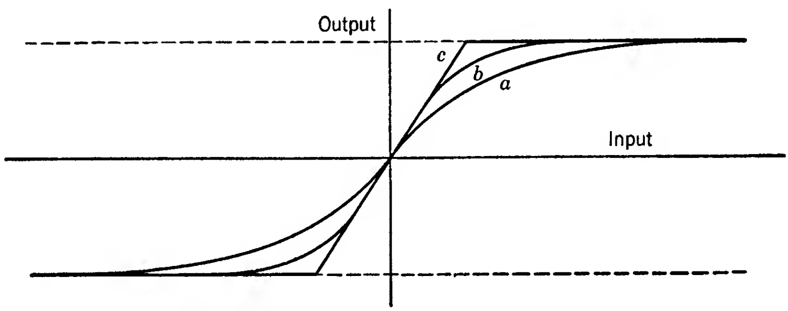


Fig. 12.23. Input-output characteristics with feedback.

ponents and cross products between frequency components that may not have been present in the input. Thus, there will always exist some "harmonic distortion," which is a consequence of the nonlinearity of the output-input relationship. If e_0 is small, the powers of e_0 will be even smaller and the harmonic distortion will be small. In other words, the distortion suffered by signals which are small in magnitude compared to the limits of saturation is small.

It is possible to reduce the magnitudes of the higher derivatives of eq. 12.24 in order to reduce the distortion. The most economical way to do this is with negative feedback. Consider Fig. 12.23 in which curve a represents the system without feedback. Curve b is for the system with typical amounts of feedback. Curve c is for an infinite amount of feedback. In all cases, the gain has been normalized to be the same for small signals. If the signal is less than the saturation limits, the effects of negative feedback in reducing distortion should be clear; there will be no distortion when the feedback is infinite (neglecting frequency-varying effects). However, for signals larger than the saturation limit *no amount of feedback will help,* and with practical amounts of feedback considerable distortion will result unless signal levels are maintained moderately below the saturation limits.

At low frequencies (neglecting low-frequency p-z), feedback can be quite effective if $K\beta$ is large. However, at the higher frequencies $KG\beta$ must be made small in the interests of stability; thus, feedback is not effective for high-frequency signal components, which will therefore suffer much more distortion than low-frequency components. In fact, the high-frequency feedback may be more characteristic of positive than negative feedback, with the result that the distortion given the high-frequency components of the signal can be worse than with no feedback at all. These distorted high-frequency components will beat with the

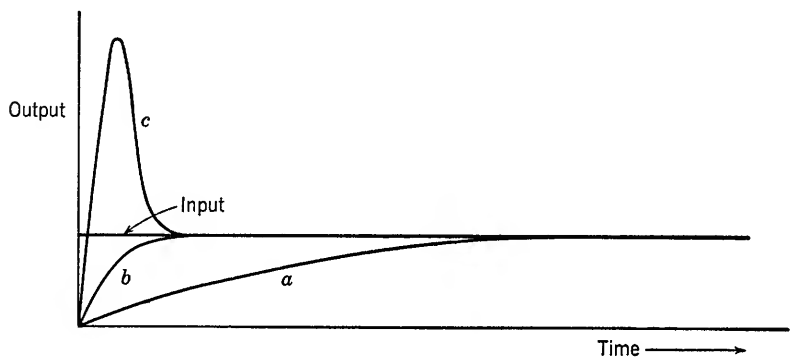


Fig. 12.24. Waveforms associated with a narrow stage.

low-frequency undistorted components to introduce spurious components which may be in the frequency range occupied by the desired signal. If amplification with low distortion is desired, the gain function of the system with feedback must drop off smoothly at high frequencies so that high-frequency components will be attenuated as much as possible. In feedback audio amplifiers, a high-frequency peak is particularly bad, often resulting in an amplifier that sounds worse with than without feedback! As a result, the closed-loop transfer function of audio amplifiers should never be made equal ripple or maximally flat; rather, linear-phase functions and even those with purely real poles are more suitable.

It was observed that negative feedback increases the bandwidth of a (possibly degraded) amplifier. The saturation limits also have a rather profound effect on this. Let us take as an example an amplifier which, without feedback, has some given bandwidth. To this amplifier is applied a step function. Let us also assume that the bandwidth is determined by one particular stage which has been shunted to ground with a capacitor (which can represent a capacitive load or ponderable electromechanical output device). The output from this stage will have the appearance of curve a of Fig. 12.24. Now suppose feedback is added

around the entire amplifier. The linear theory predicts the bandwidth with feedback will be increased over that without feedback. Then the output from this particular stage must be faster (if the system is linear), as indicated by curve b. This particular stage is within the feedback loop; consequently, the only way its output can be made faster in spite of capacitance loading is to momentarily overdrive the input. The input time function to this stage must have an appearance something like that of curve c of Fig. 12.24 in order that the output from the stage be like that of curve b.

For small inputs to the feedback amplifier, the peak of curve c may be moderate. However, for larger inputs, even though the desired output waveform is within the saturation limits, the peak of curve c of Fig. 12.24 may be far above the saturation limit. In other words, the peak of curve c may be flattened, resulting in distortion and a smaller effective over-all bandwidth than exists when the signal level is small.

The severity of an effect such as that depicted by Fig. 12.24 is proportional to the ratio of typical signal levels to saturation limits and consequently will be most severe if the open-loop bandwidth limitation occurs in the last stage or stages of a feedback amplifier. If the designer has any choice, he should degrade the bandwidth of a feedback amplifier in the earlier stages of the system. Of course, he has no choice if the degradation is furnished by a ponderable load.

Frequently, phenomena such as those described make it wise to use output stages having much larger power-handling capabilities than those inferred by average output signal levels if only small amounts of distortion can be tolerated. However, the output stage will not be nearly as bulky as that required in a system giving the same amount of output with the same distortion but without feedback.

12.7 Precision design methods

Let us examine the simple feedback system in detail to see what can be done in the way of precision design. The methods to be described are general. They will be developed by means of numerical examples.

Figure 12.25 shows the simple system again. It will be assumed that e_1 is the input, e_3 is the output, and β is a real number. The ruling equations are

$$\begin{aligned} \frac{e_3}{e_2} &= -KG = -K\frac{A}{B} \\ \frac{e_3}{e_1} &= \frac{-KG}{1 + KG\beta} = \frac{-KA}{B + KA\beta} \end{aligned} \tag{12.25}$$

where A and B are polynomials in p.

Let us consider only the high-frequency behavior by neglecting all low-frequency p-z. Typical uncompensated amplifier functions will have a number of poles (say n) because of stray capacitance and other factors. The poles of KG cause lagging phase shift and reduce the gain at high frequencies. It is the lagging phase shift that creates stability problems. Therefore, *it is best for the compensated amplifier to have no more poles than the uncompensated amplifier.*

For simplicity and because it appears to be a good technique, let us stipulate that the closed-loop transfer function have only poles. This means that the function A in eqs. 12.25 is unity and that the open-loop

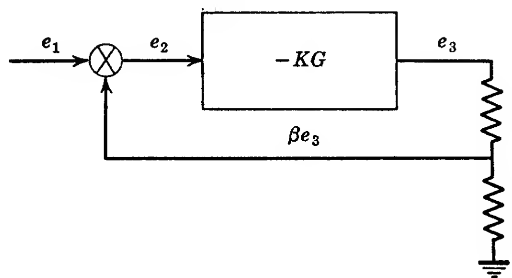


Fig. 12.25. The simple feedback amplifier.

function as well as the closed-loop function has only poles. Typical uncompensated amplifiers do not have zeros (ignoring coupling and bypass networks). Whatever compensating networks we decide to use may have zeros, but only if the zeros are cancelled by poles from other interstage networks in the amplifier. (The introduction of zeros that are not cancelled will be described later.)

Thus we have the open- and closed-loop transfer functions reduced to

$$\frac{e_3}{e_2} = -KG = \frac{-Kb_0}{p^n + b_{n-1}p^{n-1} + \cdots + b_1 p + b_0} \tag{12.26}$$

$$\frac{e_3}{e_1} = \frac{-Kb_0}{p^n + b_{n-1}p^{n-1} + \cdots + b_2 p^2 + b_1 p + b_0(1 + K\beta)} \tag{12.27}$$

where the open-loop transfer function e_3/e_2 at $\omega = 0$ is $-K$, as has been our convention.

Equation 12.27 gives the closed-loop behavior as an n-pole (no zero) transfer function. For any given $K\beta$, we may adjust the coefficients $b_0, b_1, b_2, \cdots$ so that the function is maximally flat, equal ripple, or some other prescribed function, with whatever bandwidth appears reasonable. Once we have determined these coefficients, then the open-loop transfer

function is determined. In particular, the locations of the poles of the open-loop function are determined. These data can then be used to determine suitable amplifier interstage networks so that the open-loop function will be as required.

As an example, let us study a three-pole amplifier made up, for example, from two or three grounded-cathode tubes. Let us require the closed-loop function to have a third-order pole on the negative real axis. This does not give a very sophisticated transfer function, but it may be quite justifiable for a high-fidelity audio amplifier or a pulse amplifier having a step-function response without overshoot. The closed-loop bandwidth should not be larger than the bandwidth of the uncompensated amplifier, or else trouble will be encountered in the compensation.

We thus have the closed-loop transfer function

$$\frac{e_3}{e_1} = \frac{(-Ka^3)/(1 + K\beta)}{(p + a)^3} = \frac{(-Ka^3)/(1 + K\beta)}{p^3 + 3ap^2 + 3a^2p + a^3} \tag{12.28}$$

Comparing eqs. 12.28 and 12.27, the required open-loop transfer function is seen to be

$$\frac{e_3}{e_2} = \frac{(-Ka^3)/(1 + K\beta)}{p^3 + 3ap^2 + 3a^2p + a^3/(1 + K\beta)} \tag{12.29}$$

If $K\beta$ is large, the coefficient of p^0 in eq. 12.29 will be small. This results in the open-loop function having one pole near the origin, which amounts to bandwidth degradation.

Let us normalize to frequency by setting $a = 1$. The closed-loop half-power bandwidth is therefore $(2^{1/3} - 1)^{1/2} = 0.51$ (bandwidth narrowing). To be specific, let us set $K\beta = 100$ (which is a large amount of feedback). The open-loop function becomes

$$\frac{e_3}{e_2} = \frac{-0.0099K}{p^3 + 3p^2 + 3p + 0.0099} \tag{12.30}$$

which can be factored with numerical methods to give

$$\frac{e_3}{e_2} = \frac{-0.0099K}{(p + 0.00331)(p^2 + 2.997p + 2.99)} \tag{12.31}$$

which has poles at

$$\begin{aligned} p_1 &= -0.00331 \\ p_2,\ p_2^* &= -1.5 \pm j0.863 \end{aligned} \tag{12.32}$$

The closed-loop function is shown in Fig. 12.26*a*. The corresponding open-loop function is shown in Fig. 12.26*b*, where the pole near the origin gives bandwidth degradation and where the complex poles indicate the need for a resonant-type interstage circuit for optimum compensation of the amplifier. The pole locations in Fig. 12.26*b* are typical of the results of precision design.

Suppose we had wanted the three-pole system to have a maximally flat closed-loop transfer function with a bandwidth B normalized to

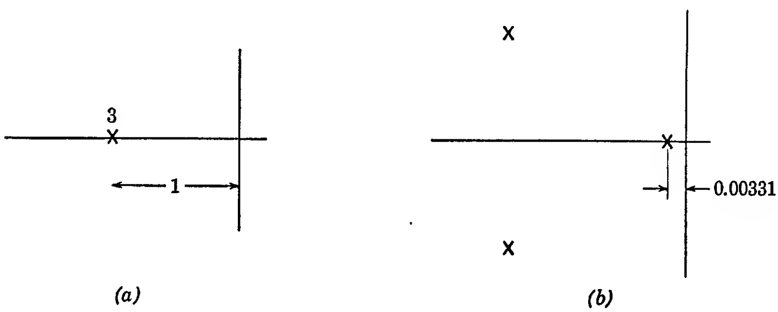


Fig. 12.26. Open- and closed-loop functions.

unity. Again, we shall use $K\beta = 100$. The closed- and open-loop transfer functions are

$$\frac{e_3}{e_1} = \frac{-0.0099K}{p^3 + 2p^2 + 2p + 1}$$
$$\frac{e_3}{e_2} = \frac{-0.0099K}{p^3 + 2p^2 + 2p + 0.0099} \qquad (12.33)$$

The poles of the open-loop function are at

$$p_1 = -0.00497$$
$$p_2, p_2^* = -0.998 \pm j0.996 \qquad (12.34)$$

which have positions similar to those of the previous example, even though the closed-loop transfer function is quite different. Note, however, that the wider band closed-loop function has indicated somewhat less bandwidth degradation.

Finally, let us consider a four-pole system. In order to have a numerical example, we shall take $K\beta = 100$ as before and a closed-loop transfer function having a fourth-order pole on the negative real axis at $p = -1$.

The closed- and open-loop transfer functions are

$$\frac{e_3}{e_1} = \frac{-0.0099K}{(p+1)^4} = \frac{-0.0099K}{p^4 + 4p^3 + 6p^2 + 4p + 1}$$
$$\frac{e_3}{e_2} = \frac{-0.0099K}{p^4 + 4p^3 + 6p^2 + 4p + 0.0099} \tag{12.35}$$

With numerical factoring, we find the poles of the open-loop transfer function as

$$\begin{aligned} p_1 &= -0.00248 \\ p_2 &= -2 \\ p_3, p_3{}^* &= -1 \pm j1 \end{aligned} \tag{12.36}$$

which has the same basic characteristics as those displayed by the three-pole function.

So far, we have discussed only the case for real β. Quite commonly, β is complex, most frequently having the form

$$\beta = \beta_0 \frac{ap + 1}{bp + 1} \tag{12.37}$$

where it is almost always desirable to have $a \geq b$. For $a = b$, β is purely real, which is the situation already discussed.

For β given by eq. 12.37, the closed-loop transfer function can be found from eq. 12.27 as

$$\frac{e_3}{e_1} = \frac{-Kb_0(bp + 1)}{(p^n + b_{n-1}p^{n-1} + \cdots + b_1 p)(bp + 1) + b_0[(b + K\beta_0 a)p + (1 + K\beta_0)]} \tag{12.38}$$

and the open-loop function is given by eq. 12.26 as before.

Whenever a and b are not equal, the closed-loop transfer function has a zero which cannot be cancelled with a pole. The zero position is determined entirely by the β circuit. If a rectangular plot of the denominator of eq. 12.38 is set up, it will be observed that for $b > a$ stability is worsened as compared to β purely real. On the other hand, for $a < b$, stability is improved because the coefficient of the first power of p is increased.

Figure 12.27 shows the standard feedback circuit modified with capacitors. A simple analysis of the β circuit of this figure gives

$$\beta = \beta_0 \frac{ap+1}{bp+1} = \frac{R_2}{R_1+R_2}\left\{\frac{pR_1C_1+1}{\{[R_1R_2(C_1+C_2)]/(R_1+R_2)\}p+1}\right\} \tag{12.39}$$

which can be adjusted to give either $a > b$ or $b > a$.

Frequently, stray capacitance makes it impossible to neglect C_2. It is possible to add a small C_1 in parallel with R_1 to make β purely real.

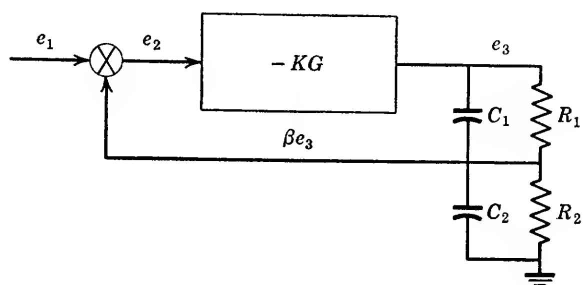


Fig. 12.27. Feedback circuit with complex β.

From eq. 12.39, the condition yielding $\beta = \beta_0$ in spite of capacitance is found by equating the pole and zero positions as

$$R_1C_1 = R_2C_2 \tag{12.40}$$

If C_2 is finite and C_1 is not added to the circuit, stabilization may be considerably more difficult.

In order to tie down some of the theory expounded so far to a practical example, consider the circuit of Fig. 12.28. The feedback is from the output to the cathode of tube V_1 and is controlled by the size of the feedback resistor R. Actually, instead of adding the input e_1 and the feedback voltage, we are subtracting them; however, because subtraction is but a special case of addition, the circuit of Fig. 12.28 falls into the category of circuits we have described. This amplifier has a transformer output. The polarity at the output can be made either positive or negative by simply reversing the connections to the transformer secondary; if the connections are wrong, the feedback will be positive rather than negative and the circuit will oscillate.

Let us assume that the transformer has a large enough step-down ratio and a purely resistive load so that we may neglect the secondary stray capacitance. Also, assume that the plate current of tube V_1 is relatively small so that we can neglect any loading on the output by the feedback

circuit. The amplifier of Fig. 12.28 can then be made to correspond to the p-z of Fig. 12.26. The one pole near the origin is furnished by the R-C plate load on tube V_1. The pair of complex-conjugate poles is furnished by the output circuit of tube V_2. If incidental resistance, secondary capacitance, and magnetizing inductance are all neglected, the

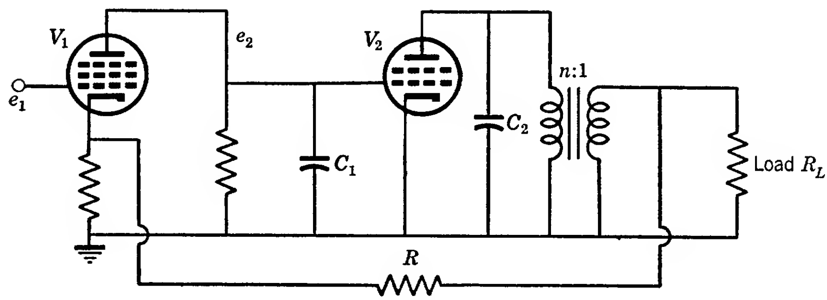


Fig. 12.28. A feedback amplifier.

equivalent output circuit, referred to the primary, will be as shown in Fig. 12.29, where n is the transformer turns ratio. This equivalent can be adjusted to have a pair of complex-conjugate poles in the desired positions.

Typically, the circuit of Fig. 12.28 would have a transformer step-down ratio of about $n = 50$ (as when the load is a speaker) and a value of K of perhaps 200. Because we have set $K\beta = 100$ in previous numerical examples, we shall assume this $K\beta$ here also, which requires that

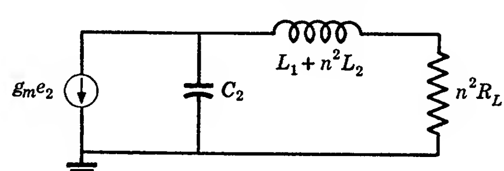


Fig. 12.29. Equivalent output circuit of the feedback amplifier.

$\beta = 100/200 = 0.5$. Accordingly, feedback resistor R will be relatively small. The gain with feedback is found from eq. 12.27 to be $K/(1 + K\beta) = 200/101 \cong 2$. Although the voltage gain is small, the power gain is considerable because the load impedance is small.

In the actual amplifier, we would be likely to by-pass the cathode of V_2 and perhaps the screen of V_1 (but not usually the screen of V_2, which is often returned directly to the plate supply voltage). Also, we would use an R-C coupling network between V_1 and V_2. The mag-

netizing inductance of the transformer and the load resistor provide another coupling network function. Therefore, the amplifier has effectively two coupling networks, a cathode by-pass circuit, and a screen by-pass circuit. The low-frequency p-z collection is therefore as shown in Fig. 12.30. If the R-C coupling time constant is made large, the open-loop low-frequency cutoff will be determined primarily by the magnetizing inductance of the transformer, as in Fig. 12.30*a*, and the system will be stable. Because of saturation problems related to a degraded stage (see Sec. 12.6), it is probably better to make R-C small and use a trans-

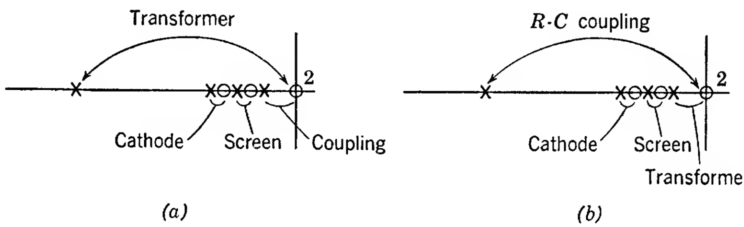


Fig. 12.30. Low-frequency p-z of the feedback amplifier.

former with a large magnetizing inductance so that the low-frequency cutoff is determined by the R-C coupling network, as in Fig. 12.30*b*.

A small capacitor in shunt with feedback resistor R of Fig. 12.28 overcomes capacitance from the cathode of V_1 to ground (which includes cathode-to-filament capacitance) as described before.

The amplifier we have been discussing has poles as given by Fig. 12.26, which are normalized to frequency. Suppose the amplifier is an audio amplifier and it is desired to obtain the *full* benefits of feedback up to about 5000 cps. Of course, feedback will provide some benefits to much higher frequencies. Accordingly, the bandwidth of the degraded stage must be about 5000 cps. Since the poles of the output stage are about 300 times as far from the origin of Fig. 12.26*b* as the pole of the degraded stage, the natural bandwidth of the transformer should be roughly 300×5 or 1.5 mcs. Thus the transformer leakage reactance should be quite small. It has already been pointed out that the magnetizing inductance should be large. In essence, a high-quality transformer is necessary. Lower quality units do not permit the full benefits of feedback to be obtained at frequencies as high as 5000 cps (or do not permit such large values of $K\beta$).

Several other tuning schemes for the audio amplifier of the example come to mind. In particular, if the transformer is somewhat overdamped, the transfer function of the output stage will have two poles on the negative real axis. One of these two poles can furnish bandwidth

degradation. The second pole can be cancelled with a zero obtained from the plate load of tube V_1. The pair of high-frequency complex-conjugate poles can also be furnished by the output circuit of tube V_1. If the load on V_1 is a shunt-peaked circuit modified to the extent of adding a resistance in parallel with the capacitance (in addition to the resistance in series with the inductance), the desired pole and zero locations can virtually always be realized. This particular tuning scheme is capable of realizing the poles of Fig. 12.26*b* with less expensive and more readily available transformers which do not have natural bandwidths as large as megacycles. Of course, if the output stage degrades the bandwidth, saturation problems with high-frequency signals may be encountered, as in a pulse amplifier for fast signals. However, if the bandwidth of the degraded output stage is large enough to accommodate most expected signal frequencies (as is the case in the audio amplifier when the degraded bandwidth is on the order of 5000 cps), saturation problems are minimized.

Had there been a total of three tubes in the amplifier of the example, the open-loop function would have four poles, which should be located according to eqs. 12.36. The output stage, as before, could furnish the complex-conjugate pair of poles, and the first and second stages the poles on the negative real axis. One of the plates of the first two stages should be degraded with a shunt capacitor, preferably the first, if possible, because of saturation problems. If $K\beta = 100$ for the four-pole system (as in the three-pole system) and $K = 10{,}000$ (which is typical), β becomes $100/10{,}000 = 0.01$ and the closed-loop gain at low frequencies becomes $10{,}000/101 = 99$. With the particular numbers used, the three- and four-pole systems reduce nonlinearities in the output stage equally. Feedback with the four-pole system also reduces nonlinearities in the stage that drives the output stage, which is not true for the two-tube amplifier. As before, a less costly output transformer can be operated somewhat overdamped to furnish real poles and p-z cancellation can be employed. The first interstage can furnish one high-frequency pole if its plate load is a simple R-C circuit. The second stage can furnish the complex-conjugate pair of high-frequency poles and one zero for cancelling purposes if its plate load is the modified shunt-peaked circuit.

So far, we have considered only all-pole closed- and open-loop functions. The addition of zeros can considerably increase the degraded open-loop bandwidth for a given $K\beta$ and for a given number of poles *in excess* of zeros (which is the only fair comparison because each interstage must furnish at least one more pole than zero due to stray capacitance). However, the arbitrary placement of zeros and subsequent analysis does

not appear to result in a facile method for introducing zeros. Therefore a slightly different approach is needed.

Consider again the Nyquist plot where the closed-loop transfer function is given by $1/\beta$ (a simple number) times the *ratio* of the two phasors, $KG\beta$ and $1 + KG\beta$. If $K\beta$ is large, the open-loop locus $KG\beta$ can be quite irregular for frequencies lower than the over-all 180-degree phase-shift frequency without resulting in a significant irregularity in the closed-loop transfer function; the ratio of the two phasors changes but little even though each phasor individually may change considerably. Since the open-loop function at these lower frequencies is dependent primarily on the bandwidth degrading function (the one low-frequency pole), it would appear that the most promising way for introducing zeros is in association with the degrading function and without modifications in the locations of any of the high-frequency poles.

With certain restraints, it should therefore be possible to utilize a "cluster" of both p-z at low frequencies as a substitution for the one degrading pole found necessary before. Of course, the closed-loop transfer function will no longer be precisely that specified and realized with simple one-pole degradation, although it may be close enough to make the difference insignificant. There are three restraints placed upon the cluster of p-z. The first is obvious; that is, the gain of the cluster function at $\omega = 0$ must be the same as that of the single-pole function in order that the open-loop gain K will be unchanged. The second restriction is that the cluster must always have one (and only one) more pole than zero. This insures that the high-frequency asymptotic behavior of the cluster will be the same as that of the simple one-pole function. This second restriction allows the degrading function to be obtained with any reasonable input impedance having some shunt capacitance. The third restraint, which is not too specific, is that the high-frequency asymptotic behavior of the cluster must be reached before the phase shift contributed by the high-frequency poles becomes too large.

The irregularity at low frequencies because of the cluster must not be excessive if the closed-loop transfer function is to be essentially the same as with the simple one-pole bandwidth degrading function. From the Nyquist plot, it should be apparent that larger values of $K\beta$ permit more irregularity. The greater the irregularity, the greater can be made the open-loop degraded bandwidth. Therefore larger values of $K\beta$ need not always require smaller open-loop bandwidths, as seemed necessary with one-pole degradation.

As a specific example and for a normalized one-pole degraded bandwidth of 1 radian, let us study the replacement of the one-pole function

with another function (the shunt-peaked function) having one zero at $z_1 = -2$ and two poles at $p_1, p_1^* = -1 \pm j1$. This function obeys all the restraints listed (with the proper constant multiplier). It is a low-pass type of function with a slight peak at a frequency somewhat below its half-power frequency. The bandwidth of the simple one-pole function is unity (by normalization). The bandwidth of the new function is 1.8; therefore, an 80-per-cent increase has been achieved in the degraded open-loop bandwidth. If more irregularity can be tolerated (that is, a larger peak), a greater advantage is possible.

Let us return now to the two-stage audio amplifier with transformer output used as an example before. The open-loop function now has one pair of high-frequency complex-conjugate poles (in the same positions as before because we have modified only the degrading stage), one pair of low-frequency complex-conjugate poles, and one low-frequency zero. The transformer output stage can furnish the pair of low-frequency poles quite easily; however, the natural bandwidth of the transformer is now $5000(2)^{1/2} = 7070$ cps and, by itself, has a two-pole maximally flat transfer function. (The bandwidth of the transformer is not increased by as much as 80 per cent because the transfer function of the transformer does not include the low-frequency zero.) The high-frequency poles and the low-frequency zero can easily be realized with the modified shunt-peaked circuit as the plate load of the first stage.

Degrading functions containing several p-z can realize quite considerable benefits in increasing the degraded bandwidth. What results is a function similar to that found with Bode's log-frequency procedure (see references for Chap. 12), except that more control is exercised here on the high-frequency characteristics. When several p-z are employed, the Nyquist plot may extend far into the region where the phase shift is greater than 90 degrees (but safely less than 180 degrees) before the open-loop gain falls to a low value.

The equation applicable to the simple one-pole bandwidth degrading function is

$$\frac{1}{p + \alpha} \tag{12.41}$$

where the single pole is at $p = -\alpha$. A degrading function with n poles and having the correct low- and high-frequency behavior is

$$\frac{p^{n-1} + a_{n-2}p^{n-2} + \cdots + a_1 p + a_0}{p^n + b_{n-1}p^{n-1} + \cdots + b_1 p + a_0\alpha} \tag{12.42}$$

For a given α, it is evident that the degraded open-loop bandwidth may be made many times larger than α, subject only to the restraint

that the asymptotic behavior is reached before the total open-loop phase shift becomes too large. If $a_{n-2} = b_{n-1}$ in eq. 12.42, the frequency where the asymptotic behavior is reached is minimized. Thus it is probably wise to make the sum of the pole and zero distances of the cluster function approximately equal, although this is not necessary if the asymptotic behavior is reached before the high-frequency p-z contribute an appreciable phase shift.

In the specific example given before, the one-pole degrading function was normalized to $1/(p+1)$ and the specific cluster function was $(p+2)/(p^2+2p+2)$. A somewhat greater degraded bandwidth can be obtained by making the cluster function more irregular, as $(p+2)/(p^2+p+2)$, which slightly increases the frequency where the asymptotic behavior is reached. A much larger degraded bandwidth can be realized by moving the p-z of the cluster function further from the origin and/or by using more p-z in the cluster function. However, in order that eq. 12.42 be satisfied, it may not be possible to have the same relative p-z positions. For example, in place of the cluster function $(p+2)/(p^2+2p+2)$, we might use $(p+4)/(p^2+4p+4)$. Because eq. 12.42 must be satisfied, the new cluster function has a different p-z arrangement and the degraded bandwidth has not nearly been doubled even though the zero distance has been doubled.

In general, it will be found that as the degraded bandwidth is made larger and larger relative to that applicable to one-pole degradation, the cluster function must have more and more p-z. In the last analysis, this means that, for a given $K\beta$ and for given high-frequency p-z distances, the larger the degraded bandwidth is made, the more the system will cost because more p-z must be used in the cluster.

If the total open-loop phase shift where the cluster function reaches its asymptotic behavior is appreciably larger than 90 degrees, the approximation of the one-pole function with the cluster function (which is supposed to keep the closed-loop transfer function unchanged) begins to break down and a high-frequency peak in the closed-loop transfer function may develop. In such a case, it may be necessary to decrease the feedback β somewhat below its idealized value.

12.8 Analog computer amplifiers

A special case of the simple feedback circuit of Fig. 12.25 is the analog computer amplifier. However, the addition of the input and feedback signals is done in a somewhat different manner. The circuit is shown in Fig. 12.31, where e_1 is the input, Z' is an unavoidable impedance (usually stray capacitance possibly in parallel with a grid-leak resistor), Z_L is the

external load on the system, and R_s is the open-loop output impedance of the amplifier.

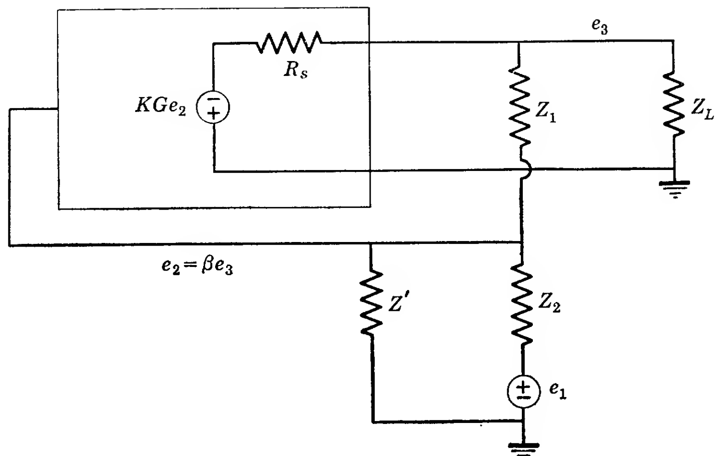


Fig. 12.31. Analog computer amplifier circuit.

The two node equations required for the solution of the circuit of Fig. 12.31 are

$$\begin{aligned} (Y_2 + Y' + Y_1)e_2 - Y_1 e_3 &= Y_2 e_1 \\ -(Y_1 - G_s KG)e_2 + (Y_1 + Y_L + G_s)e_3 &= 0 \end{aligned} \tag{12.43}$$

from which

$$\frac{e_3}{e_1} = -\frac{Y_2}{Y_1}\left[\frac{(G_s KG - Y_1)}{[(Y_1 + Y_2 + Y')(Y_1 + Y_L + G_s)/Y_1] + (G_s KG - Y_1)}\right] \tag{12.44}$$

At low frequencies and for large K, eq. 12.44 is very nearly equal to the ratio $-Y_2/Y_1$.

Let us restrict our interest to the simple analog inverting amplifier where Y_1 and Y_2 are conductances, although we shall allow for a small capacitance in shunt with G_1. The unavoidable admittance Y' will be assumed to consist of resistance and capacitance in parallel. The load

will be assumed somewhat capacitive as well. Accordingly

$$
\begin{aligned}
Y_1 &= G_1 + pC_1 \\
Y_2 &= G_2 \\
Y' &= G' + pC' \\
Y_L &= G_L + pC_L
\end{aligned}
\tag{12.45}
$$

Capacitance C_1 will be adjusted to give a real feedback factor β in spite of the unavoidable capacitance C', which requires

$$R_1C_1 = \frac{R_2R'}{R_2 + R'}C' \tag{12.46}$$

which results in

$$\frac{Y_1 + Y_2 + Y'}{Y_1} = \frac{G_1 + G_2 + G'}{G_1} \tag{12.47}$$

Finally, the open-loop no-load transfer function KG will be assumed to have only poles as

$$KG = \frac{K}{B} = \frac{K}{1 + b_1p + b_2p^2 + \cdots + b_np^n} \tag{12.48}$$

Substituting eqs. 12.45, 12.47, and 12.48 in eq. 12.44, we obtain the closed-loop transfer function as

$$
\begin{aligned}
\frac{e_3}{e_1} = &-\frac{G_2}{G_1 + pC_1} \\
&\times \left[\frac{G_sK - B(G_1 + pC_1)}{\left\{\begin{array}{l}\{(G_1 + G_2 + G')[(G_1 + G_L + G_s) + p(C_1 + C_L)]\}/G_1 \\ \quad + [G_sK - B(G_1 + pC_1)]\end{array}\right\}}\right]
\end{aligned}
\tag{12.49}
$$

Normally, G_1 and C_1 are small and the open-loop no-load gain K is large. For all ordinary frequencies, it is thus often permissible to neglect the factor $B(G_1 + pC_1)$ in comparison with G_sK. Then, the closed-loop transfer function has only poles and is given by

$$\frac{e_3}{e_1} = -\frac{G_2}{G_1 + pC_1}\left(\frac{k}{k + (1 + bp)B}\right) \tag{12.50}$$

where

$$b = \frac{C_1 + C_L}{G_1 + G_L + G_s} \tag{12.51}$$

and k is the adjusted loop gain, which accounts for the load on the amplifier and the d-c feedback factor, as

$$k\left(1 + \frac{G_2 + G'}{G_1}\right)\left(1 + \frac{G_1 + G_L}{G_s}\right) = K \tag{12.52}$$

The accuracy of the amplifier at low frequencies (ideally, $e_3/e_1 = -G_2/G_1$) is perfect to the extent that $k/(1 + k)$ is unity. If k is greater than 1000, the accuracy is better than 0.1 per cent. More load G_L and larger closed-loop gain G_2/G_1 result in lower values of k.

A typical analog computer amplifier is a three-stage direct-coupled amplifier having a beam-power output stage. It also has a differential input stage in order to realize maximum d-c stability. A d-c input voltage of zero must result in a d-c output voltage of zero. A typical uncompensated amplifier is shown in Fig. 12.32. The potentiometer in the cathode of the first stage adjusts for best compensation against d-c

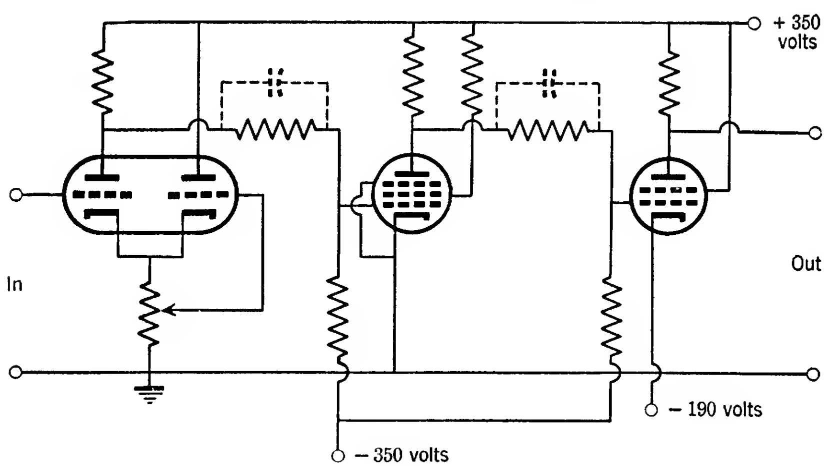


Fig. 12.32. A d-c amplifier.

drift (due to power-supply variations and tube operating point instability) and also can be used to zero the output voltage. (The grid that is returned to the cathode also affords a convenient place to inject a d-c stabilizing signal from an a-c "chopper" amplifier.)

The amplifier of Fig. 12.32 has no low-frequency poles or zeros because no by-passing or coupling circuits are employed. The low-frequency open-loop no-load gain is typically about 5000. The various plate load and divider resistances (except the load resistor of the last stage) are typically on the order of hundreds of thousands of ohms to megohms. Thus, the uncompensated amplifier has only a moderate bandwidth as limited by stray capacitances. The output resistance R_s is on the order of a few thousand ohms.

Because of the resistive divider in each interstage, and because stray capacitance appears at two places in each interstage (tube output and input capacitance), the open-loop no-load function G has four high-frequency poles. We may place small capacitors across the interstage divider resistors (shown dashed in Fig. 12.32) in order to achieve the same effects as in the β circuit, where in spite of capacitance β could be made purely real. This trick reduces the number of high-frequency poles from four to two. Including the load, it will therefore not be necessary to work with functions having more than three high-frequency poles.

Let us assume that the function G corresponding to the amplifier of Fig. 12.32 has two poles and no zeros as

$$G = \frac{1}{B} = \frac{1}{1 + b_1 p + b_2 p^2} \tag{12.53}$$

Substituting eq. 12.53 in eq. 12.50, we get

$$\frac{e_3}{e_1} = -\frac{G_2}{G_1 + pC_1}\left(\frac{k}{k + (1 + bp)(1 + b_1 p + b_2 p^2)}\right) \tag{12.54}$$

which can be adjusted as desired by suitably choosing b_1 and b_2, which determines the open-loop poles. The pole introduced by the factor $G_1 + pC_1$ is usually not important at the frequencies of interest and is neglected. In any event, it has no bearing on stability.

For simplicity, let us assume that R_s is small so that the pole due to the load can be neglected. Except in unusual circumstances, this approximation is justified. Then, eq. 12.54 becomes

$$\frac{e_3}{e_1} = -\frac{G_2}{G_1 + pC_1}\left(\frac{k/b_2}{p^2 + (b_1/b_2)p + (1 + k)/b_2}\right) \tag{12.55}$$

In order to have a specific example, let us demand that the closed-loop transfer function have a second-order pole on the negative real axis at $p = -a$. Then

$$\frac{b_1}{2b_2} = a \qquad \frac{1+k}{b_2} = a^2 \tag{12.56}$$

The open-loop no-load transfer function is therefore

$$G = \frac{a^2/(1+k)}{p^2 + 2ap + a^2/(1+k)} \tag{12.57}$$

which has poles at

$$p_1, p_2 = -a \pm a\sqrt{1 - 1/(1+k)} \cong -2a, \; -a/[2(1+k)] \tag{12.58}$$

where the approximation is valid for large k.

The undegraded stage has a pole due to stray capacitance whose position determines the closed-loop pole position at $-a$. The degraded stage is shunted with a capacitor (at the plate) so that its pole is $1/[4(1+k)]$ times as far from the origin as that of the undegraded stage. Of course, the degraded bandwidth can be increased with more sophisticated degrading functions as described in Sec. 12.7.

Problems

1. A feedback device is diagrammed in Fig. P.1. Obtain the expression for e_1/e_0. Find the positions of the p-z of e_1/e_0. What relationship must be satisfied if the gain function e_1/e_0 is to be maximally flat?

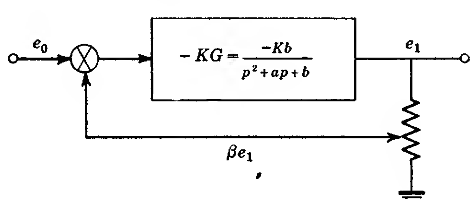


Fig. P.1.

2. If it is desired to have *exactly* $e_1/e_0 = -\beta$ for the system of Fig. P.1 at $\omega = 0$, what must be the coefficients a and b and what must be β?

3. Suppose $a = 2$ and $b = 1$ and the open-loop gain at $\omega = 0$ is 1000 in Fig. P.1. What must β be to give a fairly flat function? Sketch the closed-loop gain and phase functions for β twice that required for approximate flatness. Can this system ever be unstable?

4. It is desired to double β over that giving flatness in Prob. 3 without changing the open-loop gain at $\omega = 0$. What must the new open-loop function be to give the same closed-loop function? Sketch the p-z of KG before and after this modification.

5. Suppose the open-loop function is

$$-KG = -K\frac{p+2}{p^2+2p+1}$$

What must now be β (approximately) to give a fairly flat closed-loop transfer function? Assume $2K = 1000$.

6. If in Prob. 3 $\beta = 1$, what must be made the value of K for approximate flatness?

7. Suppose the bandwidth of Prob. 3 (closed-loop) is to be doubled without changing β or K. Sketch the required pole positions of the open-loop function. Show the p-z of two lead networks required to correct the original transfer function.

8. Suppose in Fig. P.1

$$-KG = \frac{-1000}{(p+1)^3}$$

At what value of β and at what frequency will the system begin to oscillate? Sketch the closed-loop gain function with frequency for $\beta = \frac{3}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ this critical value.

9. Suppose that a lead network is used in cascade with the function of Prob. 8 such that

$$-KG = \frac{-4000}{(p+1)^2(p+4)}$$

which has the same open-loop gain at $\omega = 0$. At what value of β and at what frequency will this system oscillate? Sketch the closed-loop gain function for β half the critical value.

10. If in the circuit of Fig. P.1

$$-KG = \frac{-1000 \times 10^{25}}{(p+10^7)^3(p+10^4)}$$

what must the value of β be in order that KG approximate a constant gain up to roughly the 180-degree phase-shift frequency? (Use the contours of constant gain.)

11. The poles at -10^7 of Prob. 10 are presumably due to R-C interstage networks. Suppose shunt-peaked interstages are used instead such that

$$\frac{10^{21}}{(p+10^7)^3} \rightarrow \frac{10^{21}(p+10^7)^3}{(p^2+10^7p+10^{14})^3}$$

Plot the p-z of the new open-loop transfer function and determine β such that the closed-loop gain will approximate a constant in the pass band. Compare the results to those of Prob. 10.

12. A lag network is used in cascade with the open-loop function of Prob. 10 having the form

$$\frac{10^{-2}(p+10^7)}{(p+10^5)}$$

Plot the p-z locations of the open-loop function and determine β for approximately constant gain in the pass band. Compare to the results of Probs. 10 and 11.

13. An open-loop function relating to a mechanical regulating device is

$$-K \frac{0.1}{(p+1)(p+0.1)}$$

The feedback factor β is unity. What must K be in order that the system be approximately flat with feedback and what is the bandwidth with feedback?

14. In cascade with the function of Prob. 13 is placed a lead network with the transfer function

$$\frac{2(p+0.1)}{p+0.2}$$

What must K be in order that the system be approximately flat and what is the bandwidth with feedback? Compare with the results of Prob. 13.

15. An open-loop transfer function is

$$\frac{-10{,}000 \times 10^{21}}{(p+10^7)^3}$$

and the feedback factor β is complex having capacitance feedback such that

$$\beta = \beta_0 \left(\frac{1+10^{-3}p}{1+10^{-7}p}\right)$$

What β_0 (at $\omega = 0$) is required to have a fairly constant closed-loop gain and what is the resulting bandwidth? Compare to that when β is constant at β_0.

16. An amplifier has an open-loop gain of 500 and an output impedance of 5K. For a β such that the closed-loop gain at $\omega = 0$ is 100, find the output impedance at $\omega = 0$. What is the approximate output impedance at frequencies near the 180-degree phase-shift point? If a gain of 100 is obtained in the amplifier before an internal disturbance is introduced, what will be the magnitude of the disturbance at the output at $\omega = 0$ and at frequencies near the 180-degree phase-shift point?

17. The amplifier of Fig. P.17 is to be used as a band-pass regenerative amplifier. Assuming $r_p \ll R$, $e_k = e_0$, $e_1 = -\mu(e_v - e_k)$, and $R_L \gg R$, determine the center fre-

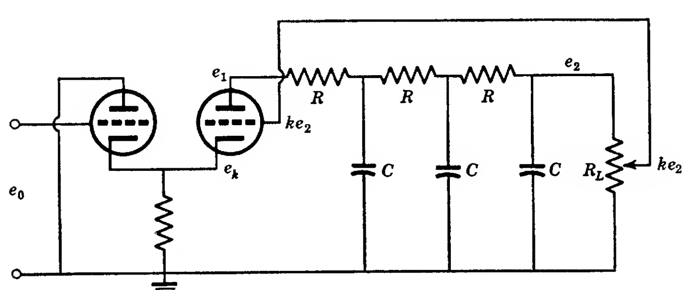


Fig. P.17.

quency of the transfer function e_1/e_0 and the bandwidth in terms of μ, R, C, and the potentiometer fractional setting k.

18. Three coupling networks with $R = 100\text{K}$ and $C = 0.01\ \mu\text{f}$ exist in a feedback amplifier. If the open-loop gain at mid-band frequencies is 1000, what must be β in order that the gain with feedback will have only a slight peak at low frequencies? What is the low-frequency half-power point?

19. If one of the three capacitors of Prob. 18 is reduced to $0.001\ \mu\text{f}$, what do β and the half-power point become?

20. A three-stage feedback amplifier has an open-loop gain of 1000. The three cathode by-pass and coupling networks give an open-loop transfer function

$$\left(\frac{p+50}{p+100}\right)^3 \left(\frac{p}{p+100}\right)^3$$

What maximum value of β can be used if a very slight low-frequency peak is the most that can be tolerated?

21. Repeat Prob. 20 if one of the three coupling networks is changed to $p/(p + 500)$.

22. The amplifier of Fig. P.22 has $C = 18\ \mu\mu\text{f}$ and $R_2 = 2\text{K}$. The tubes have $g_m = 4000$ micromhos and $K\beta = 19$. The closed-loop transfer function is to have a

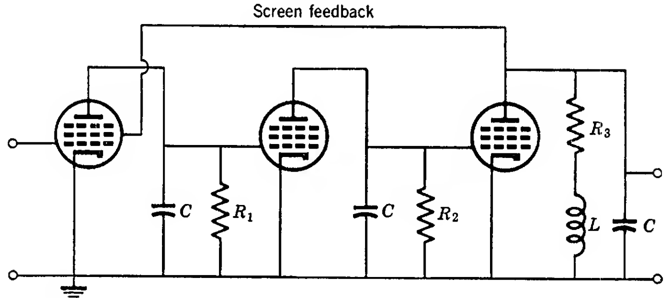


Fig. P.22.

third-order pole on the negative real axis. Determine R_1, R_3, and L. What are the open-loop pole positions? What is the closed-loop bandwidth? What is the closed-loop gain? Note that the first stage is degraded by using a large plate load resistance.

23. Repeat Prob. 22 to get a closed-loop three-pole maximally flat transfer function. Compare results to Prob. 22.

24. Repeat Prob. 22 to get a closed-loop three-pole Q-D transfer function. Compare results to Prob. 22.

25. Suppose β for the amplifier of Prob. 22 is found to be 1.2 times as large as assumed in the design. Plot the actual magnitude of the closed-loop transfer function and compare to the desired function.

26. A wide-band ten-stage amplifier having identical R-C stages has an over-all gain of 40,000 and a stage bandwidth of 20 mcs. One of the stages is degraded with a shunt capacitor in order that unity feedback ($\beta = 1$) can be used. What must be the pole position of this degraded stage so that the amplifier with feedback will be marginally stable? Sketch the closed-loop transfer function for a shunt capacitance twice the critical value.

27. Feedback is applied to the amplifier of Prob. 26 without degrading one stage. What is the critical value of β? Sketch the magnitude of the closed-loop transfer function for β half the critical value.

28. A 30-mcs band-pass amplifier has a fairly flat over-all transfer function defined by several poles. The over-all voltage gain is 10,000 and the input terminal is at an impedance level of 1000 Ω. What size capacitance between output and input will cause the system to oscillate? If the impedance level of the input is reduced to 50 Ω, what size capacitance between input and output will result in oscillation?

29. A three-pole amplifier has $K\beta = 99$. Find and plot the open-loop pole positions in order that the closed-loop transfer function be maximally flat with a (normalized) closed-loop bandwidth of unity.

30. A shunt-peaked function with a second-order pole at $p = -a$ and a zero at $p = -2a$ is used in place of one-pole degradation for the amplifier of Prob. 29. Find the parameter a. Plot the resulting closed-loop gain function of frequency and compare with that of Prob. 29.

31. Repeat Prob. 30 using a more complicated degrading function; a second-order pole at $-a$, a simple pole at $-3a$, and simple zeros at $-2a$ and $-4a$.

32. Determine the output e_4 in terms of the two inputs e_1 and e_2 for the feedback amplifier of Fig. P.32, assuming the amplifier itself (as represented by the box) has

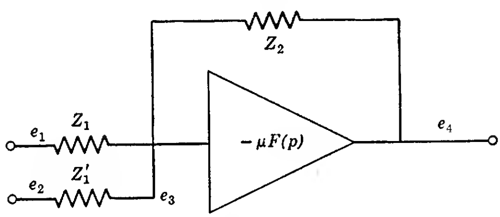


Fig. P.32.

an infinite input impedance and a zero output impedance. Also, obtain approximations valid for very large μ. Do this for the following:

a. $Z_1 = R_1$, $Z_1' = R_1'$, $Z_2 = R_2$ (summing amplifier).

b. $Z_1 = R_1$, $Z_1' = R_1'$, $Z_2 = 1/pC_2$ (integrating amplifier).

c. $Y_1 = pC_1$, $Y_1' = pC_1'$, $Z_2 = R_2$ (differentiating amplifier).

33. The amplifier of Fig. P.32 has $Z_1 = R_1$, $Z_1' = \infty$, and $Z_2 = R_2$. The gain μ is real and $F(p)$ is a three-pole function having a value of unity at $\omega = 0$. Two of the poles are at $-a$ and one of them is at $-b$, where $a > b$. What must be b in terms of a, μ, R_1, and R_2 in order that the amplifier be stable?

34. Repeat Prob. 33 for the integrator using $Z_2 = 1/pC_2$ instead of R_2.

35. Repeat Prob. 33 for the differentiator using $Z_1 = 1/pC_1$ instead of R_1.

36. If in Prob. 33 $\mu = 10^4$, $R_1 = R_2 = 10^6\ \Omega$, and $a = 10^6$ radians, what must b be so that (*a*) the system will be stable and (*b*) the transfer function with feedback will be reasonably flat?

37. Repeat Prob. 36*a* using $C = 10^{-6}$ f in place of R_2.

38. Repeat Prob. 36*a* using $C = 10^{-6}$ f in place of R_1.

39. Compare the values of b for Probs. 36, 37, and 38. Discuss the relative difficulty of stabilizing the three feedback amplifiers.

40. In the differentiator of Prob. 38, a small capacitor of value 0.05 μf is placed in parallel with R_2. Determine the value of b required to keep the system stable and discuss the accuracy of the resulting computation of the derivative.

41. Using the approximation for large μ and assuming $F(p) = 1$, determine the differential equations solved by the circuits of Fig. P.41.

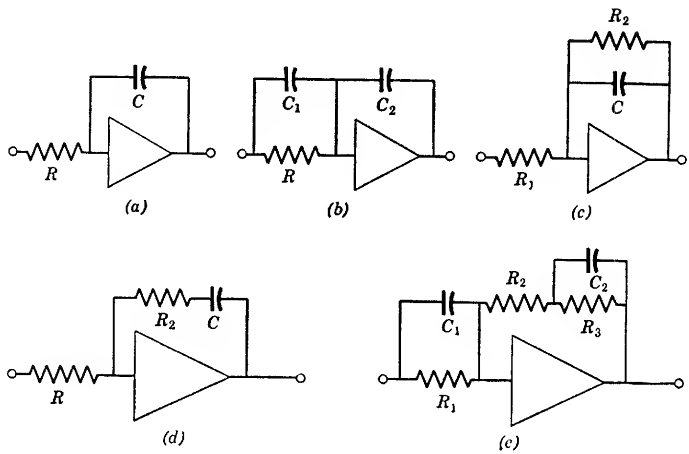


Fig. P.41.

42. Assuming large μ and $F(p) = 1$ and using only two-terminal R-C impedances for Z_1 and Z_2 in Fig. P.32, explain the permissible locations of the p-z of the transfer function. What desirable extension can be had by utilizing four-terminal networks instead of two-terminal networks?

43. Determine the p-z of the transfer functions of the circuits of Fig. P.43 and also write down the differential equations solved with the circuits. Values are in megohms and microfarads.

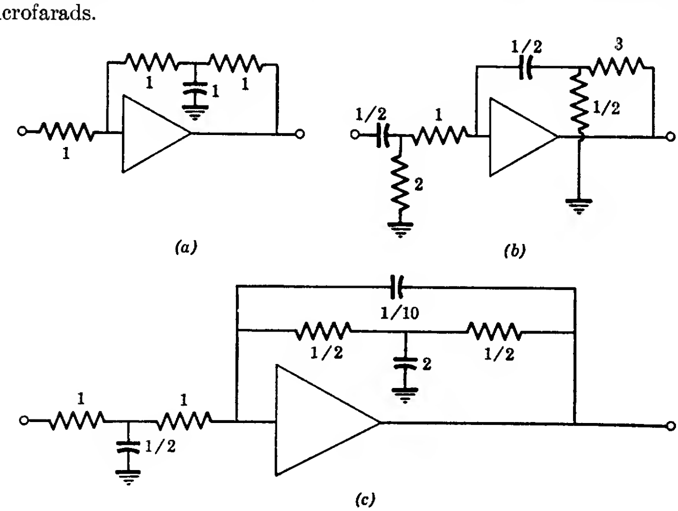


Fig. P.43.

44. If $\mu = 10^5$ and $F(p)$ has three poles with two of them at $p = -10^5$, where must the third pole be in order that the amplifier of Fig. P.43a be stable?

13

Oscillators

In the design of feedback devices, stability studies are carried out to insure that the system will operate in a condition considerably removed from that of instability so that the transfer function is suitable for the transmission of signals. In oscillators, on the other hand, design is carried out to insure that the system will *not* be stable; in other words, the transfer function *must* have poles on or to the right of the $j\omega$ axis.

There are innumerable different tube-circuit combinations that act to produce oscillations. These various devices have been invented over a period of several decades and often a type of oscillator bears the name of its inventor. Actually, there are only trivial differences between most of the circuits. The similarities will be stressed here.

In order to oscillate, the linear transfer function must show poles in the right half-plane. Thus, oscillations will build up exponentially after the system is initially energized with a speed proportional to the real part of the pole position. Obviously, the amplitude must eventually be limited by system nonlinearities. Then a linear analysis is no longer strictly valid and waveforms will no longer be perfect sinusoids. However, the linear analysis can often be extended so that a good estimate of the magnitude of oscillations can be obtained.

13.1 Theory of frequency and required gain

If the KG locus of open-loop transfer function encloses the point -1, the system will oscillate. This phenomenon is the basis of practically all oscillators. Thus, we can diagram an oscillator as shown in Fig. 13.1. The diagram is simpler than that for amplifiers because in studying conditions for oscillation no special output terminal need be considered. Further, any loading on the feedback path can always be considered to be part of $-KG$ so that

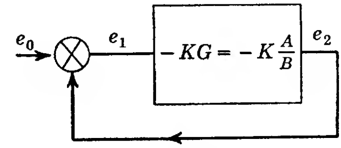


Fig. 13.1. Block diagram of an oscillator.

the feedback voltage can be assumed to be applied to a point having infinite impedance. For example, a common type of oscillator is shown in Fig. 13.2, in which the capacitance C_x in part contains the input capacitance at the grid and R_x contains in part the input resistance to the grid if the grid draws current or if the frequency is so high that transit time or cathode lead inductance is important.

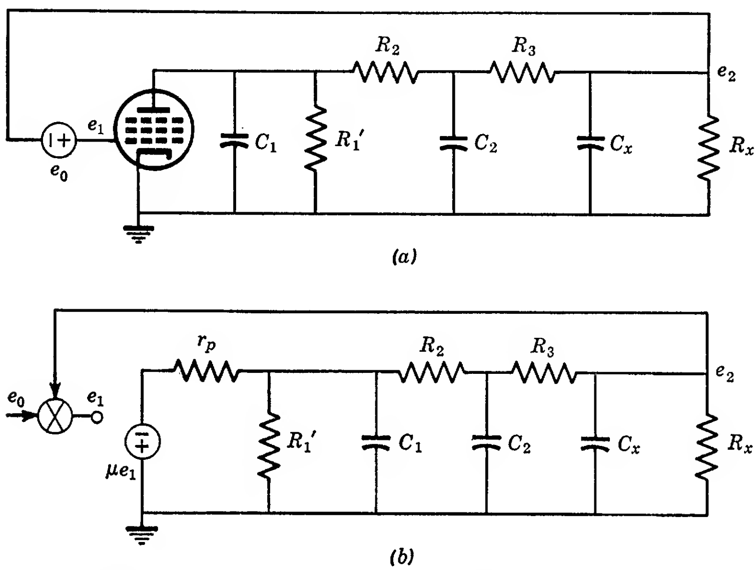


Fig. 13.2. Phase-shift oscillator employing a low-pass circuit.

The "input" can be considered to be a noise voltage (a "test" signal) applied in series with the grid, which permits the setting up of the transfer function both with and without feedback. This is not only convenient but it is a realistic way to formulate the problem because the oscillations are actually started by thermal agitation or switching transients which can be likened to noise.

Two conditions are required in order that a feedback system oscillate. First, the open-loop locus must have a shape suitable for encircling the -1 point (or $-1/K$ point if the function is normalized to gain). Second, the gain must be made large enough. The oscillation frequency will be that frequency where the phase shift is some odd multiple of 180 degrees if the tubes themselves provide an odd number of phase reversals between the input and the output of the open-loop system. The phase shift provided by the network must be an even multiple of 180 degrees otherwise.

Consider a generalized open-loop transfer function as given by eq. 13.1

$$\frac{e_2}{e_0} = -K\frac{A}{B} = -K\frac{a_0 + a_1p + a_2p^2 + \cdots}{b_0 + b_1p + b_2p^2 + \cdots} \tag{13.1}$$

Let us manipulate the transfer function as

$$-K\frac{A}{B} = -K\frac{AB^*}{BB^*} = -K\frac{(\text{Ev } A + \text{Od } A)(\text{Ev } B^* + \text{Od } B^*)}{BB^*}$$

$$= -K\frac{(\text{Ev } A + \text{Od } A)(\text{Ev } B - \text{Od } B)}{BB^*} \tag{13.2}$$

Collecting even and odd parts

$$-K\frac{A}{B} = -K \times$$

$$\frac{(\text{Ev } A \text{ Ev } B - \text{Od } A \text{ Od } B) + (\text{Od } A \text{ Ev } B - \text{Ev } A \text{ Od } B)}{BB^*} \tag{13.3}$$

The function BB^* for $p = j\omega$ is purely real. The first quantity in parenthesis in eq. 13.3 will also be purely real for $p = j\omega$, whereas the second quantity in parenthesis will always be purely imaginary. Thus we can form the tangent of the phase angle of the transfer function as

$$j \tan \theta = \left(\frac{\text{Od } A \text{ Ev } B - \text{Ev } A \text{ Od } B}{\text{Ev } A \text{ Ev } B - \text{Od } A \text{ Od } B}\right)_{p=j\omega} \tag{13.4}$$

The phase angle will be some multiple of 180 degrees when the numerator of eq. 13.4 with $p = j\omega$ is zero as

$$\text{Od } A \text{ Ev } B - \text{Ev } A \text{ Od } B = 0 \tag{13.5}$$

The solution of this equation yields in general several frequencies at which the phase shift is even as well as odd multiples of 180 degrees. The system may not be capable of oscillating at all of these frequencies; however, a simple inspection is usually adequate to determine at which frequency or frequencies the system can be made to oscillate.

At a frequency ω_i where eq. 13.5 is satisfied, the open-loop gain must be equal to or greater than unity. If eq. 13.5 is placed in eq. 13.3, we get

$$\left(\frac{-K(\text{Ev } A \text{ Ev } B - \text{Od } A \text{ Od } B)}{BB^*}\right)_{p=j\omega_i} \geq 1 \tag{13.6}$$

which determines the minimum value of K required to sustain oscillations.

Many oscillator open-loop transfer functions have either no zeros or zeros only at the origin of the p plane. In this case the equations simplify greatly. The equation of an open-loop function having zeros only at the origin is given by eq. 13.7

$$-KG = -K\frac{A}{B} = -K\frac{p^r}{b_0 + b_1p + b_2p^2 + \cdots} \tag{13.7}$$

Two situations arise, depending upon whether the number of zeros at the origin r in eq. 13.7 is odd ($r = 1, 3, 5, \cdots$) or even ($r = 0, 2, 4, \cdots$).

It is to be noted from eq. 13.7 that

$$\begin{aligned} \text{Od } A &= 0, \quad r \text{ even} \\ \text{Ev } A &= 0, \quad r \text{ odd} \end{aligned} \tag{13.8}$$

Substituting eqs. 13.8 in eq. 13.5, the equations from which the oscillation frequency can be determined become

$$\begin{aligned} \text{Ev } A \text{ Od } B &= 0, \quad r \text{ even} \\ \text{Od } A \text{ Ev } B &= 0, \quad r \text{ odd} \end{aligned} \tag{13.9}$$

Substituting eq. 13.7 in these general equations and factoring out and cancelling Kp^r (zero is not a possible oscillation frequency in linear oscillators), we get the equations yielding the oscillation frequencies from

$$\begin{aligned} \text{Od } B &= b_1p + b_3p^3 + b_5p^5 + \cdots = 0, \quad r \text{ even} \\ \text{Ev } B &= b_0 + b_2p^2 + b_4p^4 + \cdots = 0, \quad r \text{ odd} \end{aligned} \tag{13.10}$$

A factorization of these polynomials yields zeros on the $j\omega$ axis which are the frequencies at which the phase shift of the open-loop transfer function is some multiple of 180 degrees.

The required gain for oscillation can be found by placing eqs. 13.7, 13.8, and 13.9 in eq. 13.6 as

$$\begin{aligned} -K\frac{\text{Ev } A}{\text{Ev } B} &= -K\frac{p^r}{\text{Ev } B} \geq 1 \quad \text{at} \quad p = j\omega_i, \quad r \text{ even} \\ -K\frac{\text{Od } A}{\text{Od } B} &= -K\frac{p^r}{\text{Od } B} \geq 1 \quad \text{at} \quad p = j\omega_i, \quad r \text{ odd} \end{aligned} \tag{13.11}$$

If the open-loop transfer function has zeros on the $j\omega$ axis at finite values of p (perhaps in addition to zeros at the origin), eqs. 13.10 and 13.11 also apply. In all cases, however, some thought must be given to whether the phase shift is an odd or even multiple of 180 degrees at the frequencies determined by eqs. 13.10.

13.2 Examples of oscillator calculations

As a first example, consider the oscillator of Fig. 13.2. Converting the voltage source to a current source, we get the open-loop circuit of

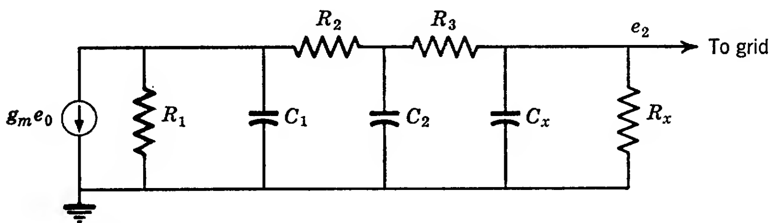


Fig. 13.3. Open-loop circuit for the phase-shift oscillator.

Fig. 13.3, in which R_1 is the parallel combination of r_p and R_1'. With three node equations, we get a transfer function with three poles and no zeros. A typical locus diagram of the open-loop transfer function is shown in Fig. 13.4. Rather than solve for the general network, we shall make assumptions to simplify the drudgery, which do not change the principles to be demonstrated in the least. Assume that

$$\begin{aligned} G_1 &= G_2 = G_3 = G \\ G_x &= 0 \\ C_1 &= C_2 = C_x = C \end{aligned} \tag{13.12}$$

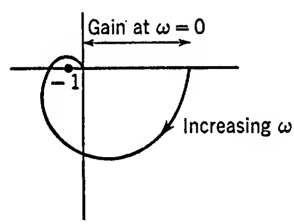


Fig. 13.4. Open-loop locus of the phase-shift oscillator.

Then, the open-loop transfer function becomes

$$\frac{e_2}{e_0} = \frac{-g_m G^2}{p^3C^3 + 5GC^2p^2 + 6G^2Cp + G^3} \tag{13.13}$$

In this case, r is even. Hence, the oscillation frequency is determined from

$$\text{Od } B = p^3C^3 + 6G^2Cp = pC^3\left(p^2 + \frac{6G^2}{C^2}\right) = 0 \tag{13.14}$$

from which

$$\omega_0 = (6)^{1/2}G/C = 2.45/RC \tag{13.15}$$

The gain at the oscillation frequency is

$$\left(-K\frac{\text{Ev } A}{\text{Ev } B}\right)_{p=j\omega_0} = \left(\frac{-g_mG^2}{5GC^2p^2 + G^3}\right)_{p=j\omega_0} = \frac{g_mR}{29} \tag{13.16}$$

In order that the system oscillate, the gain at low frequencies g_mR must be at least 29. Because of gain-bandwidth limitations in tubes,

the circuit could be made to oscillate up to about 2 mcs. When the R's and C's are not all the same, the gain required at low frequencies can be reduced considerably below 29 (minimum of 8); however, the

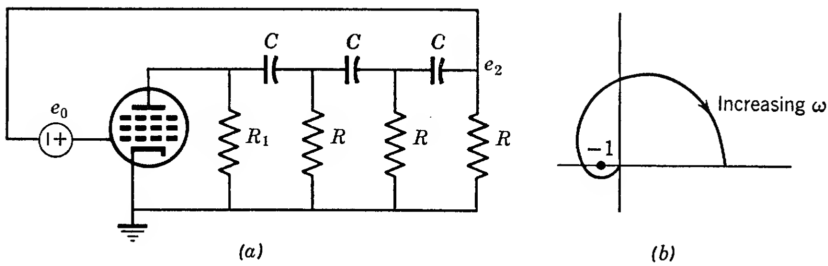


Fig. 13.5. Phase-shift oscillator employing a high-pass circuit.

practical upper limit for such oscillators is not more than a few megacycles per second.

Consider now a rather different type of oscillator useful at low frequencies where stray capacitances are not important, as shown in Fig. 13.5*a*. The open-loop locus is sketched in Fig. 13.5*b*. Let us assume that R_1 is very large (so that it can be ignored) and that the plate re-

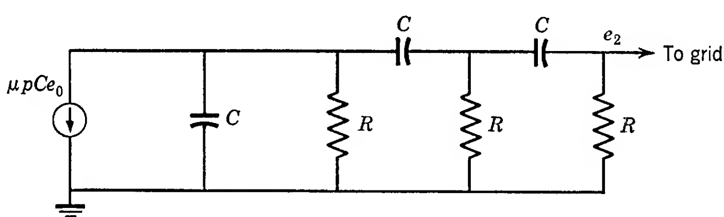


Fig. 13.6. Open-loop equivalent.

sistance of the tube is very small. Then, the equivalent circuit becomes that shown in Fig. 13.6. The transfer function is

$$\frac{e_2}{e_0} = \frac{-\mu(pC)^3}{p^3C^3 + p^2 6C^2G + p5CG^2 + G^3} \qquad (13.17)$$

This function has an odd number of zeros at the origin; hence, the oscillation frequency is found from

$$\text{Ev } B = G^3 + p^2 6C^2 G = 0 \qquad (13.18)$$

from which

$$\omega_0 = \frac{G}{(6)^{1/2} C} \qquad (13.19)$$

The gain at this frequency is

$$\left(-K\frac{p^r}{\text{Od } B}\right)_{p=j\omega_0} = \left(\frac{-\mu(pC)^3}{p^3C^3 + p5CG^2}\right)_{p=j\omega_0} = \frac{\mu}{29} \tag{13.20}$$

Thus, the amplification factor of the tube must be at least 29. With a nonuniform filter, the required amplification factor may be as small as 8.

13.3 The various types of oscillators

Consider the oscillator depicted in Fig. 13.7*a*. The equivalent circuit is shown in Fig. 13.7*b*. This is essentially the circuit of the "Colpitts"

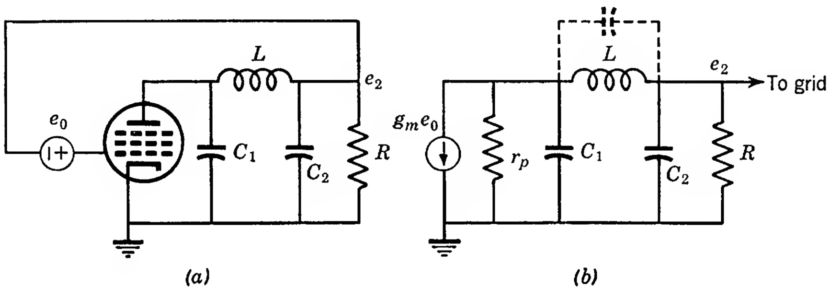


Fig. 13.7. An oscillator employing an *L-C* pi network.

oscillator more commonly seen in diagrammatic form as in Fig. 13.8. If the ground node is assumed to be the cathode (which may not actually be the case in the physical device but can be assumed so for the purpose of analysis), the circuit of Fig. 13.8 can be redrawn as in Fig. 13.7*a*.

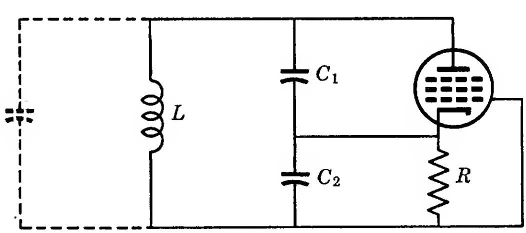


Fig. 13.8. The Colpitts oscillator.

The locus plot of the open-loop transfer function can have two different appearances, depending on the value of R, as shown in Fig. 13.9. For small R, the gain drops off with frequency more or less uniformly and is roughly similar to that of the first example of the preceding section. When the gain function in the open-loop varies relatively slowly

with frequency, the oscillator is said to be of the "phase-shift" type. Both of the examples used in the preceding section fall into this category as well as the present example for small R.

On the other hand, when the gain function has the appearance of a band-pass function as in the Colpitts oscillator with large R, the oscillator can be said to be of the "resonant" type. It must be emphasized, however, that there is no essential difference between the equations and

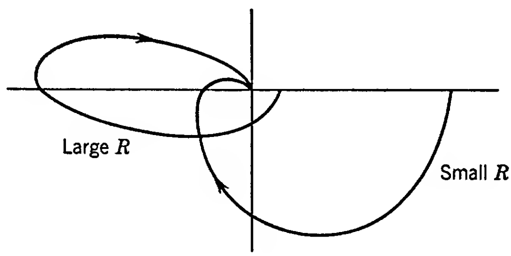


Fig. 13.9. Open-loop loci for the Colpitts oscillator.

basic requirements for oscillation in either type of oscillator. The phase-shift type makes use of a low-pass or a high-pass network, whereas the resonant type makes use of a band-pass (or rejection) network. The upper frequency limit for oscillators of the phase-shift type can thus be expected to be dependent upon the tube gain-bandwidth product, whereas the upper frequency limit of the resonant type is dependent on

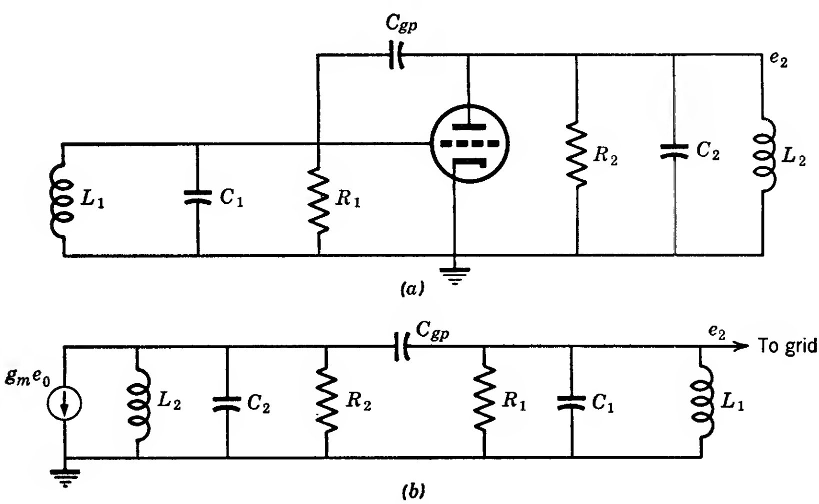


Fig. 13.10. The tuned-plate tuned-grid oscillator.

the magnitude and bandwidth of the resonant peak of the transfer function which, at very high frequencies, is ultimately determined by transit time and cathode lead inductance in the tube.

As a further example of a resonant oscillator, consider the tuned-plate tuned-grid oscillator of Fig. 13.10. The circuit is a double-tuned capacitance-coupled network with the coupling furnished by the capacitance between grid and plate of the tube. A typical open-loop locus plot for the "TPTG" oscillator is shown in Fig. 13.11. The open-loop transfer function has three zeros at the origin and two pairs of complex-conjugate poles. (If a plate-to-grid inductor is employed rather than a capacitor, the device becomes a Hartley oscillator and the filter structure is the double-tuned mutually coupled circuit.)

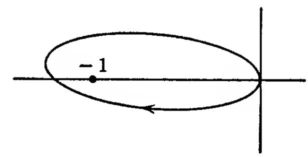


Fig. 13.11. Open-loop locus of the TPTG oscillator.

Of course, oscillators with two or more tubes can also be of value; however, unless some special feature is desired such as linearity of fre-

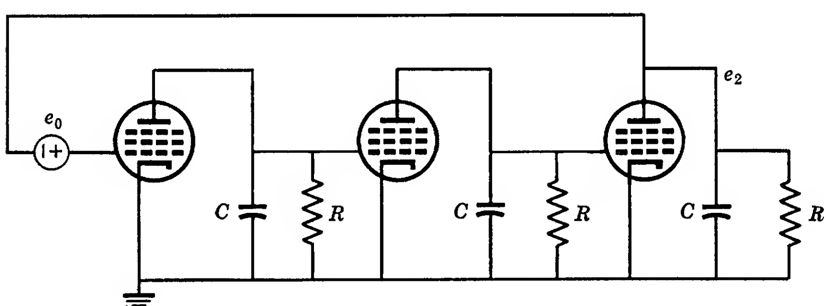


Fig. 13.12. A three-stage phase-shift oscillator.

quency tuning or large tuning range, more than one tube is somewhat wasteful. As an example, consider the three-tube phase-shift type of oscillator of Fig. 13.12. Its open-loop circuit is shown in Fig. 13.13. All tubes are assumed identical and plate resistances are ignored. The

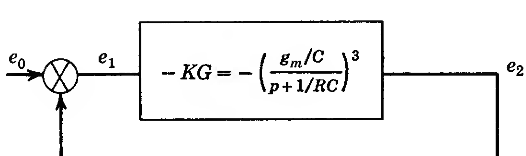


Fig. 13.13. Block diagram of the three-stage oscillator.

oscillation frequency is that where each stage produces 60 degrees of phase shift giving

$$\omega_0 = (3)^{1/2}/RC \tag{13.21}$$

The minimum gain required per stage is two; thus, such a phase-shift oscillator is capable of oscillating at a frequency as high as half the gain-bandwidth product of the tubes. Changing R and C by means of switches, it can be made to oscillate from this upper frequency to as low a frequency as is desired. A one-, two-, or three-section variable capacitor or resistor permits the frequency to be varied continuously. If $R \ll r_p$ and C is much larger than the total stray interstage capacitance, the oscillation frequency will be relatively independent of tube operating points and tube changes.

An oscillator often mistakenly considered to be a unique type is the so-called electron-coupled oscillator. This device utilizes a tetrode or pentode tube with the cathode, grid, and screen acting as a triode oscillator and with the output taken from the plate. Its advantage is that variations in the circuitry at the plate have only a second-order effect on the oscillation frequency. However, it is functionally no different than a triode oscillator followed by a pentode isolating amplifier. If the plate current in the electron-coupled oscillator is distorted from a pure sine-wave form, multiples of the oscillation frequency can be obtained from a resonant circuit in the plate tuned to the desired harmonic. In this way, the oscillator can be made to generate an output frequency higher than that of the oscillating part of the circuit. Most single-tube oscillators are not as good as the electron-coupled oscillator as harmonic generators because of the difficulty of finding a place in the circuit where a resonant circuit tuned to a harmonic frequency may be placed.

13.4 Electromechanical oscillator elements

The piezoelectric crystal has the property of developing an internal electric field when mechanically strained, and conversely, being mechanically strained when acted upon by an electric field. There exists a natural resonant frequency to the mechanical vibrations so that they become larger in amplitude when driven by an alternating electric field at the resonant frequency. The device when placed in a suitable crystal holder has an equivalent electric circuit of the series-resonant type as shown in Fig. 13.14. The capacitance C' is that furnished by the crystal in its holder (as well as any additional stray capacitance resulting from external circuitry) and is calculable as an ordinary parallel-plate capacitor. Small variations in the resonant frequency may be obtained by adding a variable capacitor in shunt with C'.

A crystal will oscillate at any one of several harmonically related frequencies. Thus, its equivalent circuit has not only a series-resonant circuit representing the lowest mode but also other series-resonant circuits representing the harmonic modes.

The main feature of the crystal is that the inductance of its equivalent circuit is extremely large (on the order of henrys at radio frequencies), whereas R and C are very small. Thus, the Q of the device is extremely

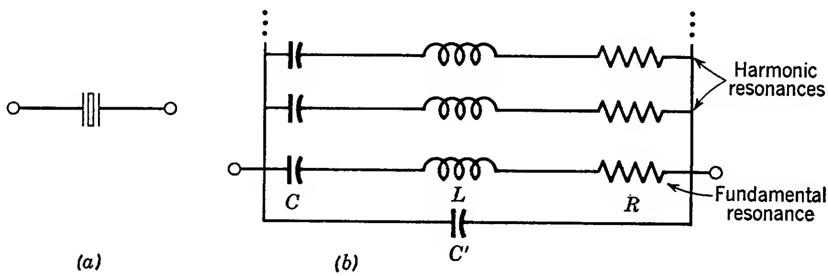


Fig. 13.14. Equivalent circuit of the crystal.

large, often as large as hundreds of thousands, which is several orders of magnitude larger than that obtainable with ordinary coils and capacitors. Thus, oscillators using crystals can be made to be extremely stable, stabilities of one part in millions not being uncommon.

Since the vibrations of the crystal are mechanical, it would be expected that the frequency range of crystals is limited. This is in fact

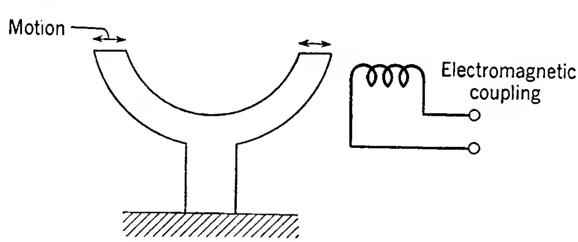


Fig. 13.15. The tuning fork.

true, crystal fundamental frequencies up to but a few megacycles per second being the practical limit. However, a crystal can be made to vibrate at harmonics of its fundamental frequency, which allows the practical upper frequency limit of crystal oscillators to be extended to something on the order of 100 mcs. If crystal-controlled frequencies in excess of this are desired, frequency-multiplying amplifiers are required.

An audio frequency equivalent to the crystal is the tuning fork. It too is a high-Q mechanically resonant circuit, although the Q is nowhere

near as high as that of the crystal (on the order of thousands). The tuning fork is made of magnetic material (at least partly) and is coupled to an electric circuit by means of a coil as in Fig. 13.15. The changes in position of the fork induce voltages in the coil, and conversely, currents in the coil create forces on the fork.

Variations in temperature have some slight effects on the natural frequency of crystals and other electromechanical resonators. Obviously, operating the device in a temperature-controlled oven will control this

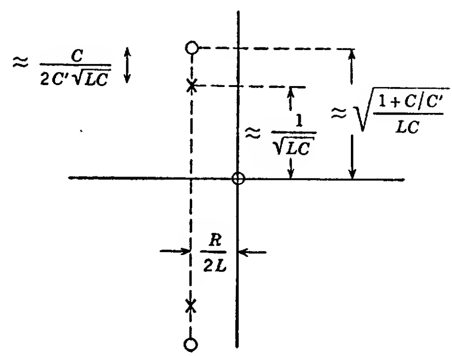


Fig. 13.16. P-z of crystal admittance.

effect. Some crystals have temperature coefficients that are practically zero over a temperature range near the ambient.

The Q of electromechanical circuits is limited by the resistance of leads carrying electric currents and also by loading provided by the air on the sides of the vibrating material. Consequently, if the maximum obtainable Q is desired, the device must be operated in an evacuated chamber. With air loading Q's on the order of thousands are common with crystals, whereas when operated in a vacuum Q's of hundreds of thousands can be obtained. (However, when a crystal is placed in a circuit, plate resistances and the like cause the circuit Q to be less than the natural Q of the crystal.)

The input admittance of the equivalent circuit for the crystal of Fig. 13.14 (neglecting all but the fundamental frequency) is

$$Y = pC' \frac{p^2 + p(R/L) + (1 + C/C')/LC}{p^2 + p(R/L) + 1/LC} \tag{13.22}$$

which has p-z as shown in Fig. 13.16. The value of $R/2L$ is so small that the imaginary part of the p-z positions is essentially no different from the radial distances of the p-z from the origin. For $C/C' \ll 1$, the

distance between the pole and the zero is very nearly $(C/2C')/(LC)^{1/2}$. Typically, this distance is considerably larger than the real part of the p-z positions.

Notice that at frequencies near the pole the crystal behaves as a series-resonant circuit with a resonant frequency relatively independent of the shunt capacitance C'. If the ultimate in oscillator stability is desired, the crystal is operated in this manner so that changes in tube and circuit capacitances that might be in shunt with C' will have little effect on the frequency.

When the crystal is operated as a series-resonant device, it is most important that it be placed in a low-impedance vacuum-tube circuit because R is small (less than a few hundred ohms). If this is not done, the Q of the resonant circuit including external circuit loading will be much less than that of the natural crystal.

If a coil is placed in parallel with the crystal (operating in a series-resonant mode) such that the added coil is resonant with capacitance C', the Q of the series-resonant circuit will be little affected. However, the Q of the parallel resonance will be much reduced so that the exact series-resonant frequency will be less dependent upon tube and circuit shunt capacitance. In fact, the parallel resonance of the crystal may be made almost unnoticeable in this manner.

At frequencies near the zero of Fig. 13.16, the crystal behaves like a parallel-resonant circuit. When the crystal is used in this mode, slight adjustments in frequency can be made by varying C'; however, the range of these adjustments is small because C/C' is always very much smaller than unity. Nevertheless, the range is adequate to permit instruments such as secondary frequency standards to be precisely adjusted. Also, a slight amount of frequency modulation can be obtained by electronically varying the capacitance in shunt with C'. To obtain maximum stability when operating a crystal as a parallel-resonant circuit, any tube or other capacitance in shunt with C' must be relatively constant and the impedance level of the circuit must be high.

Any oscillator with a series- or parallel-resonant circuit that determines to a large extent the oscillation frequency (a resonant type of oscillator) can be made into a crystal oscillator by simply replacing the resonant circuit with a crystal. The only limitations are those pertaining to the maximum frequency of oscillation and of maximum permissible power; a crystal may be damaged (physically cracked or burned) if it is forced to carry too large a current at resonance.

A crystal or other electromechanical resonant circuit can be used for filtering as well as for frequency determination in oscillators. Because of the very large values of Q obtainable with such devices, they are

useful in band-pass amplifiers. Since the crystal has closely spaced parallel- and series-resonant frequencies, it can be made to pass a very narrow band of frequencies while at the same time rejecting a frequency very near the desired frequency. All but the cheapest communications receivers are equipped with a crystal-filter circuit that can be switched into the intermediate-frequency amplifier.

13.5 Tuning methods

It is often desirable to have some mechanism for tuning oscillators. Tuning may be achieved either mechanically or electrically. Mechanical tuning is obtained by varying the value of one or more circuit elements, which adjust the frequency to some fixed value. Also, mechanical scanning at rates of a few cycles per second for sweep generator purposes can be achieved by motor-driving some tuning element. Electronic tuning causes the oscillation frequency to be proportional to a voltage, which permits the production of frequency-modulated signals.

The most common mechanical tuning method is that of varying a capacitor or several capacitors simultaneously. In a resonant-type oscillator, the frequency is usually proportional to $1/(LC)^{1/2}$ so that a 4:1 change in capacitance is required to give a 2:1 change in frequency. In an R-C phase-shift oscillator, a 4:1 change in capacitance results in a 4:1 change in frequency. Inductance tuning is another common means for changing frequency. In this, a ferrous (or sometimes copper) slug is moved in and out of a coil or sometimes a short-circuited turn is varied in position. Finally, resistance may be varied, which is particularly convenient in phase-shift oscillators at low frequencies.

Of more interest for presentation here are methods for obtaining electrical tuning. We shall mention several different schemes for doing this. First, and somewhat related to conventional mechanical methods, is the use of ferromagnetic cores for coils and ferroelectric dielectrics for capacitors. In the former, a current can be passed through a bias winding to change the permeability of the "ferrite" core and consequently change the inductance of the coil. In the latter, a bias voltage applied across a capacitor can change the dielectric constant and consequently change the capacitance. It appears that the application of ferrites to resonant-type oscillators can result in frequency tuning ranges of 2:1 or more up to frequencies of 100 mcs or so. There are, however, disadvantages to present ferroelectric and ferromagnetic materials. The first is a rather annoying temperature sensitivity and the second is a hysteresis effect that shows up in a curve of oscillation frequency versus control current or voltage.

Another electrical tuning method also analogous to mechanical tuning is obtained with controllable resistors and consequently is most ap-

plicable to phase-shift types of oscillators. For example, vacuum or crystal diodes (crystal diodes are "noisier" than vacuum diodes) can be used in place of resistors in R-C phase-shift oscillators. Diodes have the property of having a resistance that decreases as d-c control current through them is increased; consequently, a biasing current can control frequency in a suitable oscillator employing diodes. An example cap-

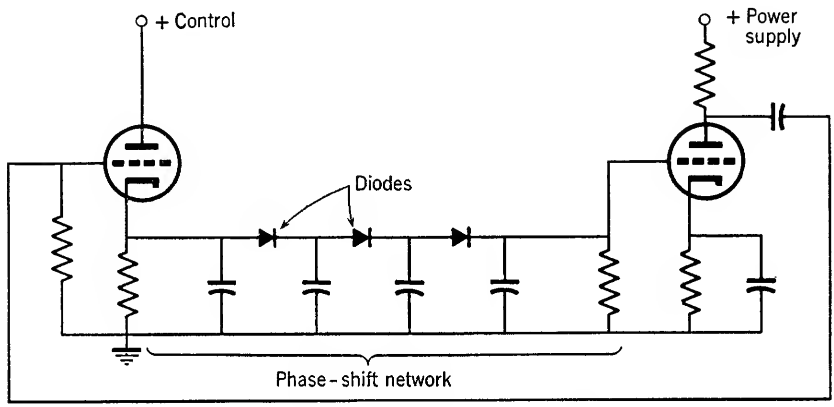


Fig. 13.17. A tunable oscillator employing diodes.

able of tuning a range of 3:1 is shown in Fig. 13.17. The main disadvantage of diode-controlled oscillators is that the oscillation magnitude must be small in order that the diode provide a large range of resistance. Another method for obtaining controlled resistance variations is with cathode followers in which the output resistance is varied according to the plate voltage (hence current).

The most common scheme for obtaining electronic tuning is not analogous to mechanical methods; it employs a "reactance tube." Essentially, a reactance tube draws a reactive current from some point in an oscillating circuit and consequently acts as a reactance which can be varied by changing the amount of current in accordance with a control voltage. Basically, the reactance modulator has the form shown in Fig. 13.18. (It should be pointed out that the reactive current need not be

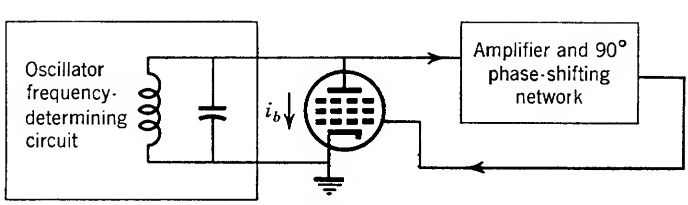


Fig. 13.18. A general reactance-tube circuit.

drawn from the same point in a circuit at which is obtained the control voltage eventually applied to the grid of the reactance tube.) A common practical embodiment is shown in Fig. 13.19, in which the R-C

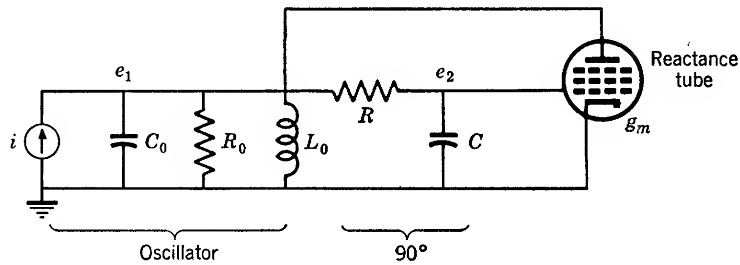


Fig. 13.19. A widely used reactance-tube modulator.

network has $\omega C \gg R$ at resonance so that e_2 lags e_1 in phase by nearly 90 degrees.

The equivalent circuit of Fig. 13.19 is shown in Fig. 13.20. The input impedance, assuming that $e_2/e_1 = 1/pRC$ and that e_2 is essentially zero

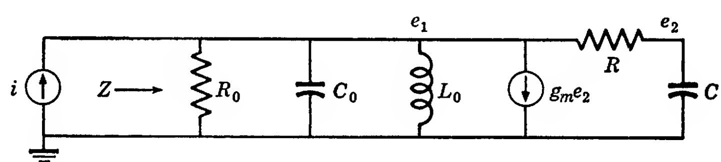


Fig. 13.20. Equivalent circuit of the widely used modulator.

as far as e_1 is concerned, is readily found from the node equations as

$$Z = \frac{e_1}{i} = \frac{p/C_0}{p^2 + p[(G_0 + G)/C_0] + [(1 + g_m L_0/RC)/L_0 C_0]} \tag{13.23}$$

If we let ω_e be the resonant frequency with $g_m = 0$ as

$$\omega_e{}^2 = 1/L_0 C_0 \tag{13.24}$$

then the resonant frequency with g_m not zero is

$$\omega_0 = \omega_e(1 + g_m L_0/RC)^{1/2} \tag{13.25}$$

The tuning range obtainable with the circuit of Fig. 13.19 is generally not more than a few per cent, primarily because the R-C phase-shift network reduces the voltage applied to the grid of the reactance tube by a fairly large amount. A greater tuning range (about 20 per cent) can be obtained with the reactance-modulated phase-shift type of

oscillator shown in Fig. 13.21. In this circuit, the phase shift across each of the two sections of the low-pass constant-k filter is 90 degrees, resulting in a reactance-tube grid voltage 90 degrees out of phase with its plate current. The current drawn by the reactance tube lags the voltage e_1 by 90 degrees and consequently acts like a variable inductance

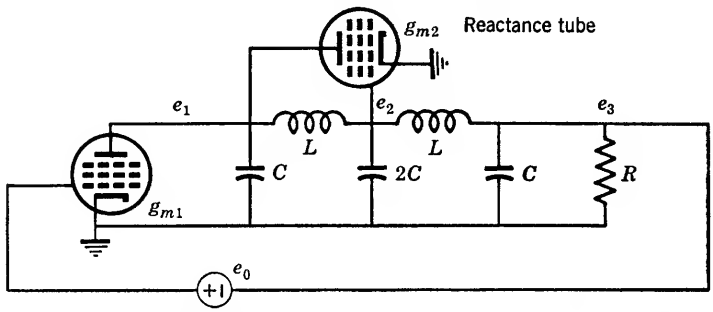


Fig. 13.21. A wide-range reactance-modulated phase-shift oscillator.

in parallel with C. The equivalent open-loop circuit for this oscillator is shown in Fig. 13.22. A nodal analysis leads to

$$e_3/e_0 = -g_{m1}/[p^5(2L^2C^3) + p^4(2L^2C^2G) + p^3(6LC^2) + p^2(4LCG + LCg_{m2}) + p(4C + LGg_{m2}) + (G + g_{m2})] \quad (13.26)$$

This circuit can oscillate at two different frequencies. With the circuit element values shown in Fig. 13.21 and with R not too large, the

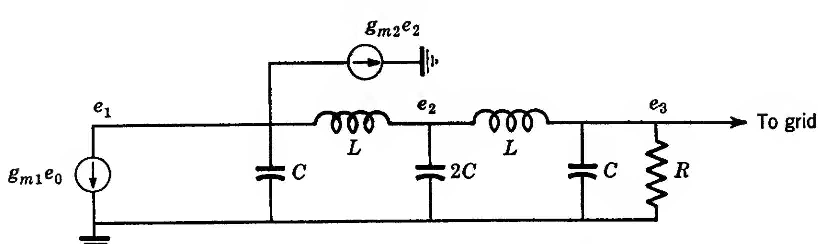


Fig. 13.22. Equivalent circuit of the wide-range oscillator.

gain required for oscillation at the lower frequency is smaller than that required at the higher frequency. Oscillations will therefore occur at the lower of the two possible frequencies. This frequency is

$$\omega_0^2 = \frac{3}{2LC}\left(1 - \frac{(1 - 2LGg_{m2}/C)^{1/2}}{3}\right) \quad (13.27)$$

The minimum frequency is given when $g_{m2} = 0$. The maximum frequency occurs when the square root in eq. 13.27 is zero, which requires that $g_{m2} = C/2LG$.

The last type of electronically tunable oscillator to be described here is the parallel-network oscillator. In this, two parallel and different networks have their outputs combined. Changing the relative gains of

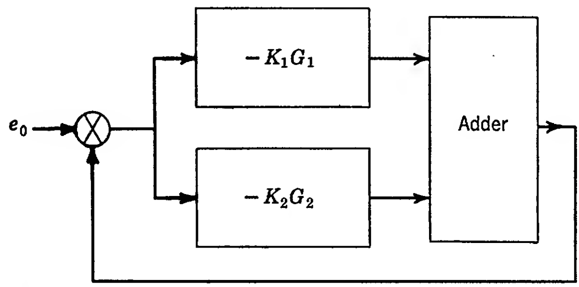


Fig. 13.23. Block diagram of a parallel-network oscillator.

the two channels (perhaps by means of transconductance variations) results in frequency tuning. The block diagram for this circuit is shown in Fig. 13.23. The open-loop transfer function is

$$\frac{e_2}{e_0} = -(K_1G_1 + K_2G_2) \tag{13.28}$$

The adding device can be either a tube or can be made of passive circuit elements.

A practical embodiment of a parallel-network oscillator is shown in Fig. 13.24. When g_{m1} is zero, oscillation will occur near frequency ω_2;

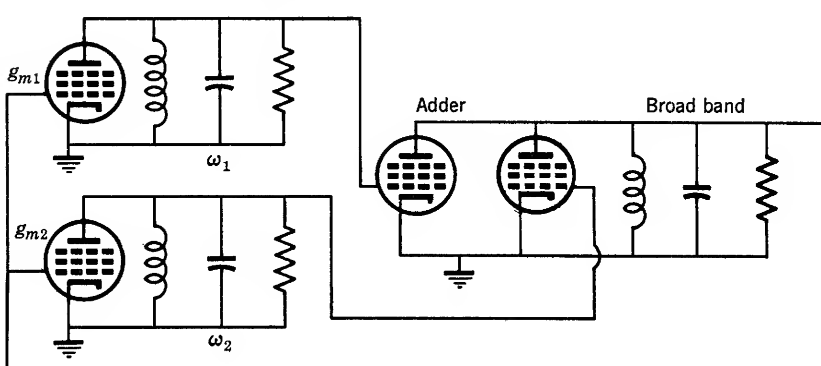


Fig. 13.24. A resonant-type parallel-network oscillator.

when g_{m2} is zero, it will occur near frequency ω_1. The variation of g_{m1} and g_{m2} can be accomplished with a balanced control voltage so that, when g_{m1} is large, g_{m2} is small and conversely. In this manner, a continuous frequency control between frequencies ω_1 and ω_2 can be obtained. The limiting tuning range of the circuit of Fig. 13.24 appears to be dependent upon the gain-bandwidth product of the tubes employed.

The parallel-network oscillator makes use of a type of network we have not discussed; that is, a composite network made up by adding the outputs of two or more separate networks, all of which are driven from the same source. If the individual networks have the same p-z, the over-all transfer function will have the same p-z as one of the networks. If the individual networks all have different p-z, the over-all transfer function has all the poles of all the networks as well as a number of zeros whose positions on the p plane can be varied by changing the relative gains of the individual networks.

13.6 Amplitude and stability of oscillations

The p-z plot of the linear closed-loop transfer function of an oscillator must show at least one pair of complex-conjugate poles to the right of the $j\omega$ axis if the gain is sufficient to cause oscillation. Usually the real part of this pole position is proportional to the gain, as is indicated in

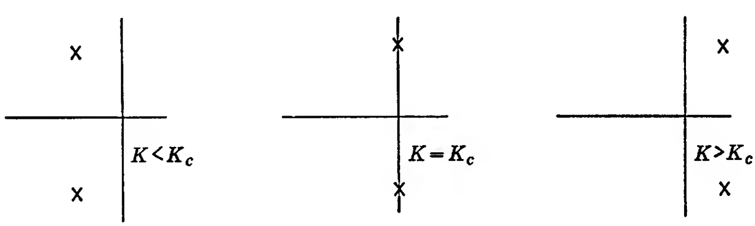


Fig. 13.25. Effect of gain on a pole position.

Fig. 13.25, where K is the actual gain and K_c is the critical gain. The gain when oscillations first build up from noise and/or switching transients must be greater than the critical gain. After oscillations have been established, the gain must not fall lower than the critical gain or else the oscillations will die out, which results in an unstable oscillator.

Now consider the input-output amplitude function of the amplifying device contained in an oscillator. It will be characterized by gradual saturation limits. When a sine wave is applied to the input, a somewhat distorted sine wave (obtained by graphical construction) will ap-

pear at the output, as indicated in Fig. 13.26. Although the output sine wave will be somewhat distorted, a first approximation is to assume that the peak value of the output wave is the peak value of the fundamental Fourier component of the wave. If we assume this, we may plot from Fig. 13.26 a new curve giving the ratio of e_{out} to e_{in} as indicated by Fig. 13.27 in which both e_{out} and e_{in} are sine waves.

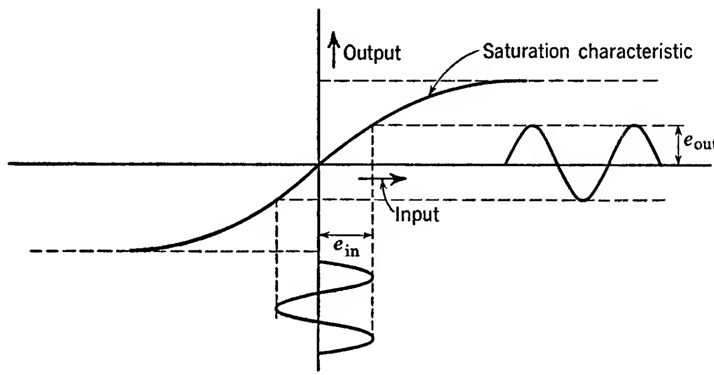


Fig. 13.26. Input-output saturation curve.

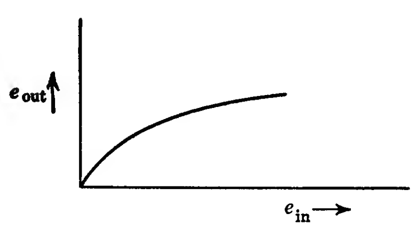


Fig. 13.27. Output sine-wave amplitude.

The gain of a device is defined as the ratio of output to input. This ratio can be obtained from the curve of Fig. 13.27 to yield a plot as indicated in Fig. 13.28. The result is a curve which generally shows that maximum gain occurs for small input signals with the gain uniformly decreasing as the input signal level increases.

When an oscillator is initially energized, signal levels are small, maximum loop gain exists, and the poles lie in the right half-plane. However, as the oscillation builds up, the loop gain is decreased, which causes the poles to move towards the $j\omega$ axis. Ultimately, the stable situation is reached when the level of oscillation has increased to the point where the gain is reduced to the critical value and the poles lie on the $j\omega$ axis. Knowing the gain

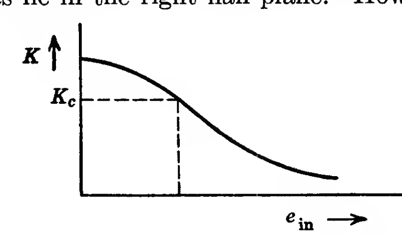


Fig. 13.28. Gain to the fundamental component.

versus input sine-wave magnitude function of the oscillator amplifying system and knowing the critical gain, a fair estimate of signal level can often be obtained.

In order to use a curve of output versus input, the load into which the nonlinear amplifying device operates must be known. This may generally be calculated from circuit element values if the oscillation frequency is known. In addition, these data, along with the graphically constructed output waveform, permit a Fourier analysis to be made in order to determine the magnitudes of the various harmonics in the output.

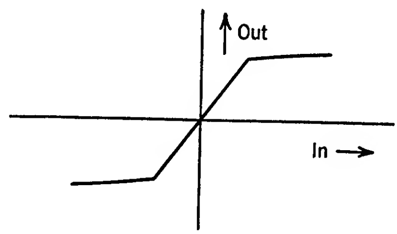


Fig. 13.29. Characteristic not favoring small oscillations.

It should be clear that the self-regulating mechanism associated with oscillators allows considerable liberty in adjusting the level of oscillation by adjusting the gain. In fact, if the output-input curve is slightly curved down to very small input levels, the oscillator can be made to oscillate very weakly. Then the pole moves but a short distance from starting to sustained oscillations and it can be expected that the frequency of oscillation will be quite independent of system gain. If the input-output characteristics have a form something like that shown in Fig. 13.29, it may not be possible to obtain weak oscillations; rather, the system may go into moderately strong oscillations abruptly.

Depending upon the circuit configuration, the frequency when oscillations start to build up may be materially different from that in the stable condition. We would then expect that changes in gain because of such things as power-supply variations and temperature would have an appreciable effect upon the frequency of oscillation as well as on the amplitude of oscillation. For this reason, automatic amplitude controlling devices are often added to oscillators when stability is important. Also, the oscillator may be operated in a temperature-controlled environment. An amplitude control system can be patterned after the automatic volume control system found in most receivers. Also, nonlinear resistors can be employed for amplitude control so that the resistance varies in accordance with the level of oscillation, which in turn controls some subsidiary feedback loop. For example, a light bulb in the cathode of a tube heats when the level of oscillation increases, with the result that its resistance increases, causing increased cathode degeneration and resulting in lower gain.

In addition to effects caused by gain variations, changes in oscillation frequency can occur because of changes in tube and circuit parameters. To a certain extent, the oscillation frequency is dependent upon tube plate resistance and interelectrode capacitances. All these parameters may vary from tube to tube and with time and applied power-supply voltages. Some oscillator circuits are more stable than others from this standpoint. These effects can be determined from a study of the linear system equations. We shall not dwell upon them here.

The output waveform of an oscillator is dependent upon two factors: the amount by which the loop gain exceeds the critical gain at the start of oscillations, and the type of frequency-dependent circuits employed. Clearly, if the loop gain is large, stable oscillations will drive the tubes further into the saturation region and thus distort and flatten the waveform considerably. In order to get the most nearly sinusoidal output possible, it is therefore necessary that the loop gain be only slightly greater than the critical gain. Also, the saturation characteristics must be relatively symmetric so that both positive and negative peaks of the generated waveform are distorted similarly. This reduces the second-harmonic component of the output leaving the third harmonic as the major component of distortion (which is much easier to remove through filtering than is the second harmonic).

Perhaps of greatest importance in realizing a near sinusoidal output, particularly when oscillations of large magnitude are desired, is the type of circuit employed. The distorted sine wave appears first at the plate of the oscillator tube, primarily as a distorted current waveform. The output from the oscillator, if obtained at the plate, will be dependent upon the plate load impedance. Clearly, if the plate load is a high-Q parallel-resonant circuit, considerable attenuation of the harmonics will occur and the plate voltage waveform will be a much less distorted sine wave than is the plate current waveform. In general, it can be expected that resonant-type oscillators (having circuit bandwidths less than those needed to support harmonics of the fundamental) are better from the standpoint of output purity than are phase-shift types. Further, high-Q resonant circuits are better than low-Q circuits. Similarly, it can be expected that phase-shift oscillators employing low-pass networks give outputs with less distortion than those with high-pass networks.

13.7 Ultrahigh-frequency equivalent circuits

As will be recalled from Chap. 7, the equations for any two-terminal pair can be written in a number of ways. Of particular interest, and with reference to Fig. 13.30, are

$$
\begin{aligned}
i_1 &= y_{11}{}^a e_1 + y_{12}{}^a e_2 \\
i_2 &= y_{21}{}^a e_1 + y_{22}{}^a e_2
\end{aligned}
\tag{13.29}
$$

$$
\begin{aligned}
e_1 &= z_{11}{}^b i_1 + z_{12}{}^b i_2 \\
e_2 &= z_{21}{}^b i_1 + z_{22}{}^b i_2
\end{aligned}
\tag{13.30}
$$

For bilateral networks, $y_{12}{}^a = y_{21}{}^a$ and $z_{12}{}^b = z_{21}{}^b$. *This equality is not true for networks containing energy sources such as vacuum tubes, transistors, and related devices.*

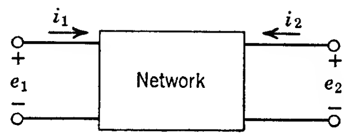


Fig. 13.30. Nomenclature for the general two-terminal pair.

It should be clear to the reader that the equivalent circuits of Figs. 13.31*a* and *b* represent eqs. 13.29 and 13.30 respectively. These equivalent circuits contain two generators. When the network has three rather than four terminals, we can replace one of the two generators with an impedance. If the impedance is selected properly, the equivalent circuits of Fig. 13.31 become those of Fig. 13.32. The reader should have no difficulty proving this equivalence.

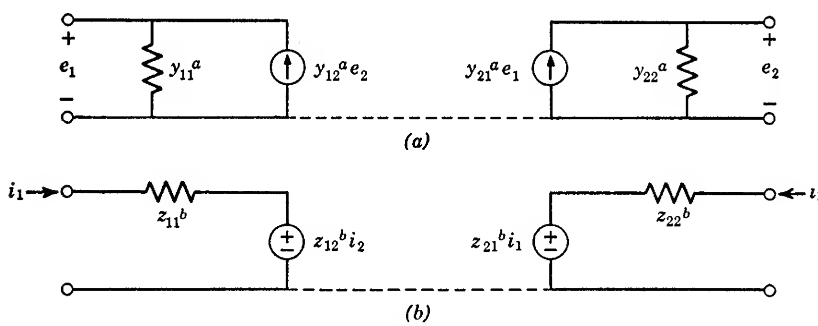


Fig. 13.31. Equivalent circuits of the two-terminal pair.

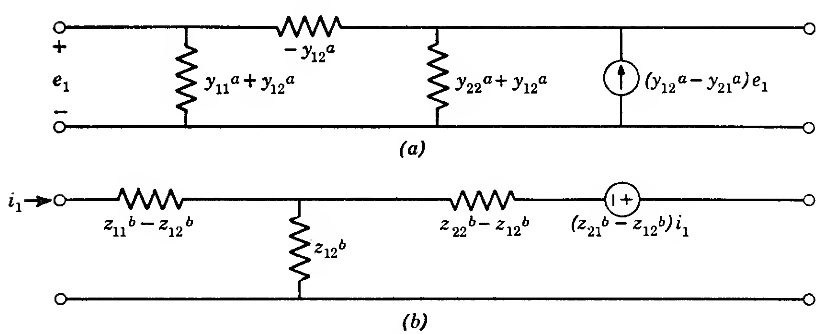


Fig. 13.32. Modified equivalent circuits containing one generator.

In particular, we can use the representations of Fig. 13.32 for just the active device itself. For example, Fig. 13.32*a* represents a negative-grid triode or a pentode with a constant screen-to-cathode voltage when the admittances are associated with stray capacitances, plate resistance, and transconductance. For triodes and pentodes, Fig. 13.32*a* leads to the equivalent circuit of Fig. 13.33.

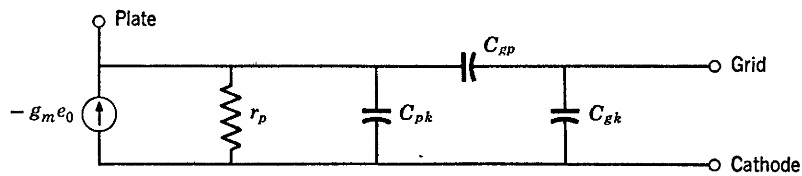


Fig. 13.33. Vacuum-tube equivalent with noise generator.

We can proceed to attach all manner of external circuit elements to the representation of Fig. 13.33. Then, if we assume that the current generator produces a noise current $-g_m e_0$ and calculate the transfer function, the open-loop oscillator transfer function is defined. It is this procedure that has been employed in earlier parts of this chapter.

At very high frequencies, the circuit of Fig. 13.33 must be modified. First, the inductances of the lead wires must be accounted for. Rather

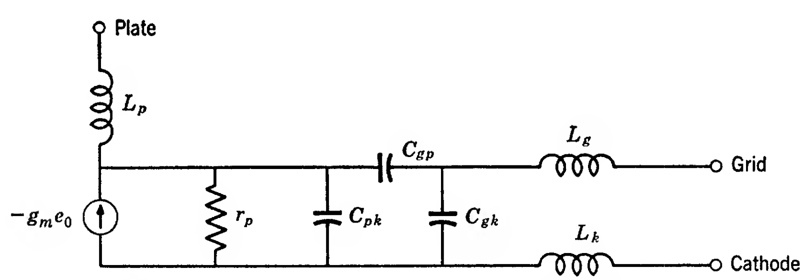


Fig. 13.34. Vacuum-tube equivalent with lead inductances.

than go back to the basic circuit of Fig. 13.32 and define new admittances, it appears more meaningful to modify Fig. 13.33 directly. The modification is shown in Fig. 13.34, where L_k is the cathode lead inductance, L_p is the plate lead inductance, and L_g is the grid lead inductance.

The next thing to be accounted for at very high frequencies is the transit time of the electrons in the tube itself. Transit time introduces two significant modifications. First is a loss at the grid, partly in the form of a conductance increasing as the square of frequency. We shall represent this phenomenon with an admittance Y in parallel with the

grid-to-cathode capacitance. In addition, the transconductance g_m is affected; its magnitude decreases with frequency and its phase becomes more lagging. To a good approximation, we can represent the current generator of Fig. 13.34 with an ideal current generator at the end of a low-pass filter.

The final high-frequency equivalent for the tube becomes that shown in Fig. 13.35. This circuit is the one that must be used in oscillator calculations at ultrahigh frequencies and at microwaves. Clearly, calculations become quite complicated, particularly at frequencies where the phase shift contributed by the filter is appreciable.

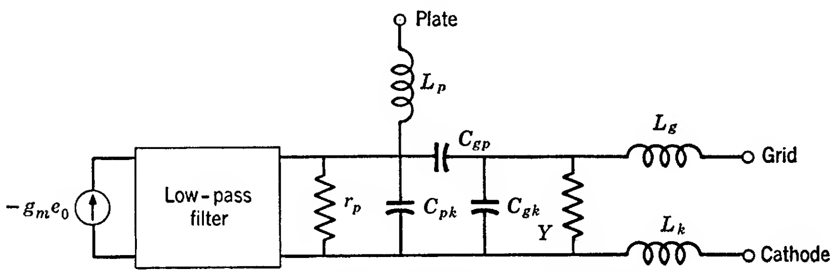


Fig. 13.35. Complete high-frequency equivalent.

Of course, the complete filter representation for transit time is almost never required. Rather, a one- or two-section approximation (either an *R-C* or an *L-C* filter) is generally adequate.

It should be noted that triode oscillators at frequencies as high as 10,000 mcs have been reported. Phase shifts due to transit time may, in some cases, amount to hundreds of degrees. In fact, it is possible to build a phase-shift type of oscillator where the required 180-degree phase shift is contributed primarily by transit time. Even at moderate frequencies, the phase shift resulting from transit time may be large enough to require a significant correction to oscillator calculations that neglect transit time.

Finally, it should be mentioned that transistor oscillators can be studied in much the same manner as has been done here. The *R-C* filter equivalent then applies to the current gain parameter a and may be important even at relatively low frequencies. With transistors, it may be more convenient to employ the T equivalent of Fig. 13.32*b* rather than the pi equivalent.

Problems

1. An open-loop transfer function has poles at -1, -2, -3, -4, and -5 and no zeros. At what frequency can an oscillator with these poles be made to oscillate if (*a*) K is negative and (*b*) K is positive?

2. Repeat Prob. 1 for a function having n zeros at the origin and n poles at -1.
3. If the function of Prob. 2 is

$$\frac{Kp^n}{(p+1)^n}$$

what minimum value of K is required at each of the possible oscillation frequencies? K may be either positive or negative.

4. An oscillator has an open-loop function

$$\frac{\pm Kp}{p^2 + 2ap + b}$$

At what frequency will it oscillate and what minimum value of K is required?

5. An open-loop transfer function is

$$\frac{\pm Kp}{(p^2 + 2ap + b)(p^2 + 2cp + d)}$$

Plot the p-z (assume the high-Q case) and show the point on the $j\omega$ axis at which oscillation will occur. Determine the minimum required value of K. Show the circuit of the oscillator.

6. Repeat Prob. 5 assuming three zeros at the origin rather than one.
7. A one-tube oscillator employs a narrow-band double-tuned mutually coupled interstage network with one Q infinite. Assuming the function is maximally flat (for convenience only), determine the effect on the oscillator frequency of a small change in the capacitance in shunt with each side of the transformer.
8. Repeat Prob. 7 using the double-tuned capacitance-coupled circuit.
9. For the Colpitts equivalent circuit of Fig. 13.7, neglecting the plate resistance, derive a relation between the minimum g_mR at the oscillation frequency and the element values, and in particular the capacitance ratio C_2/C_1.
10. Broad-band resonant circuits tuned to the same frequency are employed in a TPTG oscillator and a crystal is placed between grid and plate. The device does not oscillate without the crystal in spite of the grid-to-plate capacitance. Draw the equivalent circuit and determine the oscillation frequency relative to the center frequency of the broad-band circuits.
11. A crystal is used in place of the grid resonant circuit in a TPTG oscillator. The parallel-resonant frequencies of the lumped-element circuit and the crystal are approximately the same. At what frequency will the device oscillate?

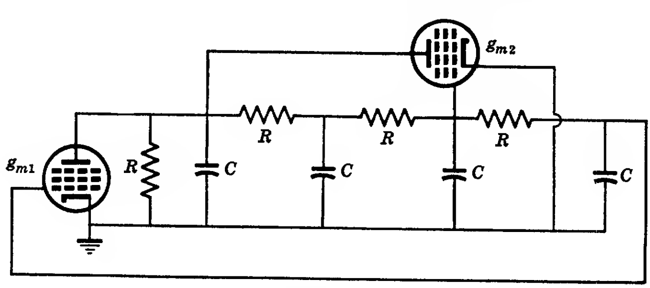


Fig. P.12.

12. A phase-shift oscillator has the circuit of Fig. P.12. Determine the oscillation frequency and the required g_mR at this frequency in terms of R, C, and g_{m2}. What is the ratio of maximum and minimum frequencies obtainable by varying g_{m2} from zero to some maximum value?

13. A parallel-network oscillator has ideal phase-shifting networks such that the open-loop gain is

$$-(K_1e^{-Tp} + K_2e^{-aTp})$$

where $K_1 = K_0(1 \pm k)$ and $K_2 = K_0(1 \mp k)$ and $0 \lesseqgtr k \lesseqgtr 1$. Determine the oscillation frequency and the frequency tuning range. Sketch the open-loop gain at the oscillation frequency as functions of k for $a = 1, 1.5, 2,$ and 3.

14. An open-loop transfer function is

$$e_2 = -K\iint e_0\,dt\,dt$$

What values of K are required for oscillation when the loop is closed and what is the oscillation frequency? What is unusual about such an oscillator? Will the system start spontaneously?

14

Servomechanism Functions

In this chapter, we shall discuss the basic properties and ruling equations of linear servomechanisms. The details of the gadgetry will be minimized; only the barest essentials of these will be presented. We shall not treat servomechanism design in the conventional manner which makes use of the Nyquist plot; rather, we shall develop methods by which the closed-loop behavior can be made a maximally flat, equal-ripple, or some other *prescribed* function. In addition, we shall discuss rate feedback, the effect of a load, and the integrating servomechanism.

14.1 The definition of a servomechanism

Before we get too far, it must be emphasized that a linear servomechanism is nothing more than a type of feedback device. However, because of its special usefulness and unique circuitry, the special title of

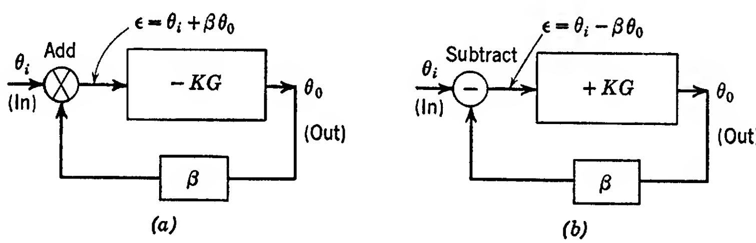


Fig. 14.1. Block diagrams for single-loop servomechanisms.

servomechanism is given. Anything that is not a servomechanism can be called a feedback amplifier or a regulator.

Let us diagram a general feedback device as in Fig. 14.1, where the open-loop function is $-KG$ for Fig. 14.1*a* and $+KG$ for Fig. 14.1*b*. The circuit of Fig. 14.1*a* is perhaps more familiar because we used it exclusively in Chap. 12. However, that of Fig. 14.1*b*, which uses a subtractor rather than an adder, is just as good a representation of a feedback device. The closed-loop functions corresponding to the cir-

cuits of Fig. 14.1 are

$$\frac{\text{Output}}{\text{Input}} = \frac{\theta_0}{\theta_i} = \begin{cases} \dfrac{-KG}{1 + KG\beta}, & \text{adder} \\[2ex] \dfrac{+KG}{1 + KG\beta}, & \text{subtractor} \end{cases} \tag{14.1}$$

from which we see that the difference between the two representations of Figs. 14.1 is quite trivial.

Let us define a servomechanism as a feedback device in which the transfer function with feedback is *precisely* $\pm 1/\beta_0$ at zero frequency, where β_0 is the value of β at $\omega = 0$. This requires that *the open-loop function $\pm KG$ have at least one pole precisely at the origin of the p plane.* In addition, *the feedback voltage must be finite at $\omega = 0$.*

The reader probably realizes the problems involved in putting a pole of the open-loop transfer function precisely at the origin of the p plane with methods we have so far studied. In fact, the only way it can be done with electronic circuits is to use a precisely controlled amount of positive feedback. Because of vacuum-tube instability, it is essentially impossible to maintain a pole precisely at the origin; rather, the pole will wander back and forth about the origin. If we desire any precision at all, this kind of positive feedback must therefore be rejected.

It must be concluded that *a servomechanism is a feedback system employing two different kinds of energy* (such as electric and mechanical or electric and hydraulic); otherwise a pole of the open-loop transfer function cannot be placed *exactly* at the origin of the p plane.

There are innumerable different types of linear servomechanisms. However, their basic properties are all similar to those of the simple servomechanism employing an electronic amplifier and a d-c motor; consequently, we shall study mainly this one type.

Any all-electronic feedback device we shall term a feedback amplifier. Any feedback device utilizing some type of nonelectric energy and one that does *not* have a pole of its open-loop transfer function at the origin of the p plane will be termed a regulator.

14.2 The motor as an output device

The motor of a d-c servomechanism is generally a shunt- or compound-wound motor with a separately excited main field. On the field winding is impressed a direct voltage so that the field in the motor is constant. The control voltage is impressed on the armature so that armature voltages of opposite polarities cause rotation in opposite directions.

If θ_m is the angular rotation of the shaft of the motor, e is a constant voltage applied to the armature, and we disregard transient disturbances for the moment, then

$$\theta_m = K_v \int e\,dt = (K_v/p)e \tag{14.2}$$

which describes a transfer function having a pole exactly at the origin of the p plane; as long as a voltage is applied to the armature, the armature rotational angle θ_m grows linearly with time. K_v is called the "velocity constant" because

$$p\theta_m = K_v e = \frac{d\theta_m}{dt} = \omega_m \tag{14.3}$$

where ω_m is the angular velocity of the armature which is proportional to the applied voltage e through the constant K_v. K_v must of course also contain conversion units between voltage and angular velocity.

If a voltage e is abruptly applied to the armature of the motor at rest, the armature will gradually come up to its final speed. The delay in achieving a constant speed is due to two factors. The first is the inertia of the rotor and the viscous frictional damping resulting from the retarding effect of bearings and windage. The second is the opposition to current change brought about by the resistance and inductance in the armature itself.

To determine the relationship between applied voltage and armature angle θ_m, it is necessary to study the electromechanical differential equations of the motor. The torque T furnished by the armature with a constant magnetic field is proportional to the armature current. This torque overcomes three opposing torques; one due to acceleration, another to frictional damping, and the last to a load torque

$$T = k_t i_a = J\frac{d^2\theta_m}{dt^2} + f\frac{d\theta_m}{dt} + T_L = (Jp^2 + fp)\theta_m + T_L \tag{14.4}$$

where J is the rotational moment of inertia of the armature and any load connected rigidly to the armature, f is the viscous friction constant, T_L is the load torque, and k_t is called the "torque constant."

It is important to limit ourselves to viscous friction where the retarding force is proportional to speed. If we wish to study the effects of Coulomb friction (static friction), we must be willing to study rather complicated nonlinear equations.

The current that flows in the armature i_a is due to an applied voltage e. A single loop equation is sufficient for determining the relation be-

tween i_a and e. Summing voltage drops, we get

$$e = i_a R_a + L_a \frac{di_a}{dt} + k_a \frac{d\theta_m}{dt} = (R_a + pL_a)i_a + pk_a\theta_m \qquad (14.5)$$

The terms i_aR_a and $L_a(di_a/dt)$ will be recognized as those pertaining to a simple R-L circuit. The term $k_a(d\theta_m/dt)$ is the back voltage of the motor, which is proportional to speed, and characterizes a d-c motor acting at least partly as a d-c generator.

If eqs. 14.4 and 14.5 are combined to eliminate the armature current i_a, the transfer function of the motor is obtained as

$$\theta_m = \frac{[k_t/(fR_a + k_ak_t)]e - \{[R_a(pL_a/R_a + 1)]/(fR_a + k_ak_t)\}T_L}{p\{[(JL_a)/(fR_a + k_ak_t)]p^2 + [(R_aJ + L_af)/(fR_a + k_ak_t)]p + 1\}} \qquad (14.6)$$

The complete transfer function of the motor is therefore seen to have two inputs, one the voltage e and the other the load torque T_L. One of the three poles is at the origin, and the other two can be found by factoring a quadratic. For $T_L = 0$, we may write

$$\frac{\theta_m}{e} = \frac{K_v}{p(T_ip + 1)(T_rp + 1)} \qquad (14.7)$$

from which the velocity constant may be identified.

The poles of eq. 14.7 usually will be real; however, if the constant k_a relating to the back voltage of the armature is quite large, they may be complex. For simplicity, we shall assume in this chapter that T_i and T_r are both real, in which case it is helpful to consider T_i as primarily the inertial time constant and T_r primarily the inductive time constant of the armature. In fact, in the limiting case of small k_a, $T_i \to J/f$ and $T_r \to L_a/R_a$. The reader should remember, however, that k_a is never zero and sometimes the poles of eq. 14.7 may be complex. Assuming them to be real does not seriously restrict the presentation here.

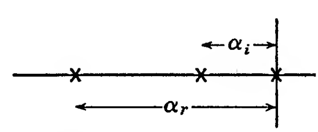


Fig. 14.2. Transfer function of a typical motor.

If we define $\alpha_i = 1/T_i$ and $\alpha_r = 1/T_r$, eq. 14.7 may be written

$$\theta_m = \frac{K_v\alpha_i\alpha_r}{p(p + \alpha_i)(p + \alpha_r)} e \qquad (14.8)$$

which has poles as shown in Fig. 14.2. Effectively, the motor acts like

a two-stage *R-C* low-pass amplifier followed by an ideal integrator. If the armature of the motor is ponderous, α_i and α_r will be small; that is, the bandwidth of the filter preceding the ideal integrator will be small. Since the bandwidth of the closed-loop system is dependent upon the bandwidth of the open-loop system, we can expect to build broader band (hence "faster") servomechanisms by using small servo motors. This, of course, is intuitively evident.

Most servomechanisms employ a d-c motor as described. Some, however, make use of a two-phase a-c induction motor. A fixed alternating voltage is applied at one of the two stator windings of the induction motor, and at the other stator winding (the control winding) an alternating voltage of the same frequency but 90 degrees out of phase is applied. The result is that the motor rotates with a torque proportional to the magnitude of the quadrature voltage. By reversing the polarity of the alternating voltage applied to the control winding (which causes it to become lagging by 90 degrees rather than leading or conversely), the motor will change its direction of rotation. Many servomechanisms utilizing a two-phase motor "chop" (that is, modulate) a direct voltage, amplify it, and apply it to the control winding. Therefore, the transfer function relating the d-c source voltage to the output angle of the motor is similar to eq. 14.7. The d-c motor is preferable in larger servos because its starting torque is larger than that of a two-phase induction motor of comparable size. However, when torque requirements are small as in instrument servos, the two-phase motor may permit a faster servo to be built because the moment of inertia of the rotor can be made quite small.

Some servomechanisms make use of a hydraulic motor and pump. They too can be expressed in the form of a transfer function such as that of eq. 14.7. The derivation of the equations, however, requires some knowledge of fluid mechanics, which cannot be presumed here.

14.3 Error detection

The block diagrams of Fig. 14.1 give no clue as to how the mechanical output is converted to a feedback voltage or how the addition or subtraction of the feedback and input is mechanized. The simplest method employs a potentiometer coupled to the output shaft of the motor and supplied with a fixed direct voltage. The circuit diagram of the input and feedback may appear as in Fig. 14.3. In Fig. 14.3*a*, an input voltage having a value between zero and V controls the output angle. For any given input e_0 and if the angle θ_m is not correct, $e_m + e_0$ will not be zero and the motor will be caused to rotate in a direction such as to make $e_m + e_0$ small, ultimately resulting in an "error voltage" ϵ of zero. In Fig. 14.3*b*, the input is a mechanical input, the only differ-

ence between Figs. 14.3*a* and *b* being that in *b* a method is indicated by which an input angular rotation is converted to an input voltage. In either Fig. 14.3*a* or *b*, as long as ϵ is small and R is large compared to the resistance of the potentiometers, the error voltage is equal to half the sum of e_0 and e_m. The two resistors R in Fig. 14.3 need not be

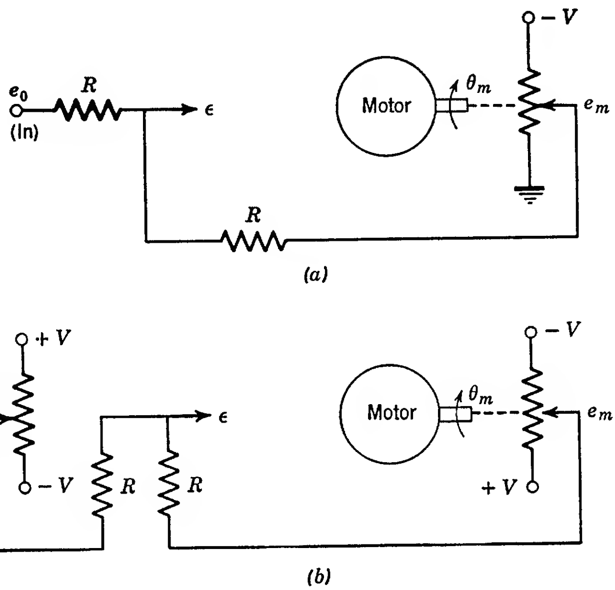


Fig. 14.3. Potentiometer error detection.

equal. In fact, these two resistors are equivalent to the resistors of an analog computer amplifier.

The more commonly used error detecting and subtracting system makes use of "synchros." A synchro system consists of a motorlike device coupled mechanically to the shaft of the servo motor and a similar device for introducing the input angle θ_i to the servo. Figure 14.4 indicates how these devices can be located. The magnitude of the

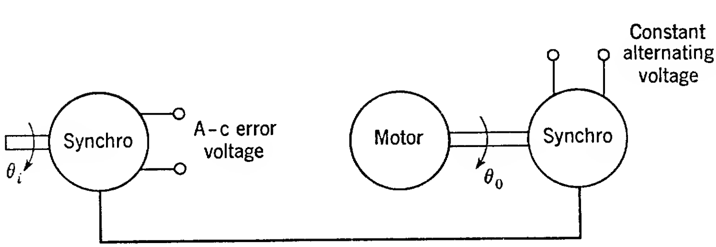


Fig. 14.4. Synchro error detection.

a-c error voltage is proportional to $\theta_i - \theta_0$, and is either in phase with the constant alternating voltage applied to the output synchro or 180 degrees out of phase with this voltage, depending upon whether θ_0 is larger or smaller than θ_i. Therefore, if the a-c error voltage is phase-detected and filtered, there results a d-c error voltage proportional to $\theta_i - \theta_0$. (It should be noted that the filtering of an error signal may introduce additional poles in the transfer function of the system.)

Any of the various error detecting and adding schemes we have discussed reduce to the same thing as far as the over-all servo behavior is concerned, which permits us to symbolize most servos as a block diagram according to Fig. 14.1. It is important to be cognizant of the various constants of proportionality relating the value of the input θ_i and output θ_0 to the error ϵ.

14.4 The simple servomechanism

Let us diagram a simple servomechanism as in Fig. 14.5 in which the error ϵ is given by the linear difference between the input and the output. This linear difference may be multiplied by a factor which includes

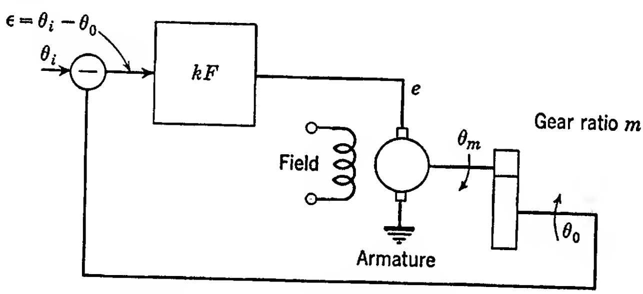


Fig. 14.5. A servo with unity feedback.

conversion units between mechanical and electrical variables (volts per degree, and so forth). However, we may always include the constant multiplier in the function kF of Fig. 14.5 so that the error ϵ can be written simply as $\theta_i - \theta_0$. The transfer function of the motor suitably modified by the gear ratio m (a torque advantage is obtained by gearing down the motor) is found from eq. 14.7 as

$$\frac{\theta_0}{e} = \frac{K_v/m}{p(T_ip + 1)(T_rp + 1)} \tag{14.9}$$

The over-all open-loop transfer function of the system of Fig. 14.5 is clearly

$$\frac{\theta_0}{\epsilon} = KG = \frac{kFK_v/m}{p(T_ip + 1)(T_rp + 1)} \tag{14.10}$$

in which the value of kFK_v/m at $\omega = 0$ is the effective velocity constant k_v. Until the integrating servomechanism is discussed, it will be assumed that F is unity at $\omega = 0$.

With $\epsilon = \theta_i - \theta_0$ in eq. 14.10, the closed-loop transfer function becomes

$$\frac{\theta_0}{\theta_i} = \frac{KG}{1 + KG} = \frac{k_vF}{p(T_ip + 1)(T_rp + 1) + k_vF} \tag{14.11}$$

For the present, let us assume that $F = 1$ at all frequencies. Then, the open-loop locus KG of eq. 14.10 appears as in Fig. 14.6. The main difference between this locus and that applicable to a typical feedback amplifier as studied in Chap. 12 is that at small ω the locus is large and has a phase angle of -90 degrees owing to the ideal integration. Because of the 90-degree phase lag given at very low frequencies, it might be expected that the stabilization of servos is more of a problem than that of feedback amplifiers. On the other hand, the servo is such a low-frequency device that it is not necessary to worry about stray capacitances with the result that the total number of poles of a servo open-loop function may be considerably less than the number associated with a multistage feedback amplifier.

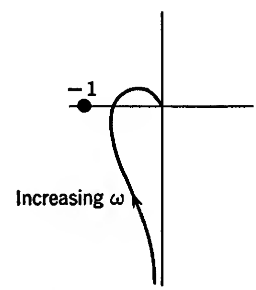


Fig. 14.6. Nyquist plot of the simple servo.

The critical gain of the system must not be exceeded if the system is to be stable. In other words, at the 180-degree phase-shift frequency, the value of the open-loop transfer function must not exceed unity (where we ignore the special cases when the locus is not similar to that of Fig. 14.6). Multiplying out the open-loop function of eq. 14.10, we get

$$\frac{\theta_0}{\epsilon} = KG = \frac{k_v}{T_iT_rp^3 + (T_i + T_r)p^2 + p} \tag{14.12}$$

The phase shift is 180 degrees when the odd part of the denominator of eq. 14.12 is zero. This frequency is

$$\omega_0 = 1/(T_iT_r)^{1/2} = (\alpha_i\alpha_r)^{1/2} \tag{14.13}$$

which is the geometric mean of the open-loop pole positions on the negative real axis of the p plane. At this frequency, the value of the

open-loop transfer function is

$$\left(\frac{\theta_0}{\epsilon}\right)_{p=j\omega_0} = \frac{-k_v T_i T_r}{T_i + T_r} \tag{14.14}$$

in which the negative sign shows it to be the 180-degree frequency. The value of eq. 14.14 must not be more negative than -1 if the system is to be stable. Therefore the critical value of k_v is

$$(k_v)_{\text{crit}} = 1/T_i + 1/T_r = \alpha_i + \alpha_r \tag{14.15}$$

which is the sum of the pole distances from the origin of the p plane. In order to have a large open-loop gain and hence a "tighter" feedback system, it is necessary that α_i and/or α_r be large so that the critical value of k_v can be made large. This requires a motor having a small moment of inertia and/or a small armature circuit time constant. For motors of even modest size, only a relatively small open-loop gain is permitted.

Let us now study the system in the closed loop. Equation 14.11 for $F = 1$ can be written

$$\frac{\theta_0}{\theta_i} = \frac{k_v \alpha_i \alpha_r}{p^3 + (\alpha_i + \alpha_r)p^2 + \alpha_i \alpha_r p + k_v \alpha_i \alpha_r} \tag{14.16}$$

which is a three-pole transfer function. The open-loop and resulting closed-loop pole positions are indicated in Fig. 14.7 for a stable system.

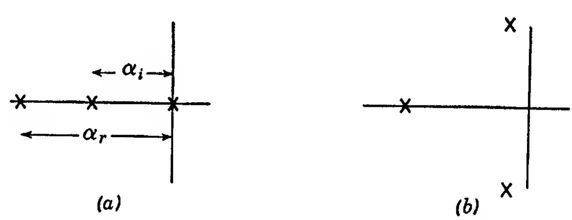


Fig. 14.7. Poles of the system with and without feedback. (a) Open loop. (b) Closed loop.

Two different situations arise: α_i and α_r approximately equal and α_i and α_r appreciably different. To study the case when they are equal, let us normalize to $\alpha_i = \alpha_r = 1$. Then the closed-loop transfer function is

$$\frac{\theta_0}{\theta_i} = \frac{k_v}{p^3 + 2p^2 + p + k_v} \tag{14.17}$$

which, no matter how k_v is adjusted (the critical value being two), gives

a rather poor (that is, nonflat) low-pass transfer function. If k_v is too large, there is a severe peak at high frequencies that does not disappear until k_v is reduced to the point where a badly sagging low-pass function is obtained.

When normalized to $\alpha_i = 1$ and for a more general α_r, eq. 14.16 is

$$\frac{\theta_0}{\theta_i} = \frac{k_v\alpha_r}{p^3 + (\alpha_r + 1)p^2 + \alpha_r p + k_v\alpha_r} \qquad (14.18)$$

which does not give a very satisfactory transfer function for α_r somewhat different from unity, although it may be adequate for some pur-

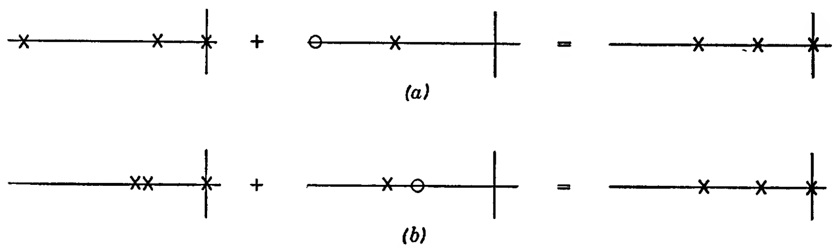


Fig. 14.8. Use of lead and lag networks. (a) α_r too large. (b) α_r too small.

poses. The best compromise appears to have α_r larger but not much larger than α_i (or conversely).

For any given α_r and α_i in eq. 14.16 or the normalized version of this equation, some k_v may be determined with trial and error procedures (the rectangular plot is convenient) which gives the most desirable closed-loop transfer function. This value of k_v will typically be approximately half the critical value.

For $\alpha_r = \alpha_i$, the closed-loop transfer function is not too good. Similarly, for $\alpha_r \ll \alpha_i$ or $\alpha_i \ll \alpha_r$, it is none too desirable. For some intermediate value where α_r and α_i are different but not greatly different, the behavior may be acceptable. Therefore it may be desirable to change the positions of the open-loop poles of the system. A lag network (for the function F of Fig. 14.5) is valuable when α_r is too large, as is indicated in Fig. 14.8a. A lead network can be used when α_r is too small, as in Fig. 14.8b. In both cases it has been assumed that α_i is unchanged. In this manner the simple servo can sometimes be compensated to a satisfactory degree.

If it is desired to increase the speed of the closed-loop system appreciably, it is necessary to increase the bandwidth of the open-loop system. This can be accomplished by moving the pole at α_i due to the armature inertia of the motor further out along the negative real axis of the p

plane and adjusting the pole at α_r to give a satisfactory closed-loop behavior. This too can be done with simple lead and lag networks. However, the only way a servo can be speeded up is to apply larger voltages to its armature. Therefore, too much "lead compensation" will result in saturation of the motor when the input to the servo θ_i changes rapidly. *No amount of lead compensation can make the motor increase its speed faster than that given at the saturation level.* We conclude that discretion is called for when attempting to speed up a servo, no matter how sophisticated the compensating networks are.

14.5 The precision design of the simple servomechanism

Maximally flat, equal-ripple, and similar functions (which give appreciable overshoot) are highly suited as servo closed-loop functions. Thus, if we can adjust the gain as well as α_r and α_i related to the open-loop function to give a maximally flat or similar closed-loop function, our design would be just about the best we could expect to achieve. Suppose, for example, we want to adjust α_r, α_i, and k_v in eq. 14.16 to give the maximally flat closed-loop function of bandwidth B as

$$\frac{\theta_0}{\theta_i} = \frac{B^3}{p^3 + 2Bp^2 + 2B^2p + B^3} \tag{14.19}$$

A moment's reflection shows us that we cannot realize the function of eq. 14.19 by adjusting the parameters of eq. 14.16. Therefore, a different approach is needed. We start by assuming an open-loop function as

$$\frac{\theta_0}{\epsilon} = \frac{k_v b}{p^3 + ap^2 + bp} \tag{14.20}$$

which leads to a closed-loop function

$$\frac{\theta_0}{\theta_i} = \frac{k_v b}{p^3 + ap^2 + bp + k_v b} \tag{14.21}$$

Now let us determine the coefficients a, b, and k_v so that eq. 14.21 will be maximally flat. By comparing coefficients of eqs. 14.19 and 14.21, we get

$$k_v b = B^3 \qquad b = 2B^2 \qquad a = 2B \tag{14.22}$$

from which

$$k_v = B/2 \tag{14.23}$$

The required open-loop transfer function is therefore

$$\frac{\theta_0}{\epsilon} = \frac{B^3}{p(p^2 + 2Bp + 2B^2)} \tag{14.24}$$

which has poles at

$$p_1 = 0$$
$$p_2, p_2^* = -B \pm jB \qquad (14.25)$$

the positions of which are shown in the p-plane sketch of Fig. 14.9. If the simple servo had an open-loop function like that of Fig. 14.9 rather than that of Fig. 14.7, adjusting the loop gain according to eq. 14.23 would automatically give the desired maximally flat closed-loop transfer function of bandwidth B. In order to get the function of Fig. 14.9, a compensating network must be placed in cascade with the open-loop function which has zeros to cancel the poles of the original open-loop characteristic and complex-conjugate poles as in Fig. 14.9. The original open-loop function and the proper compensating network function are shown in Figs. 14.10a and b respectively.

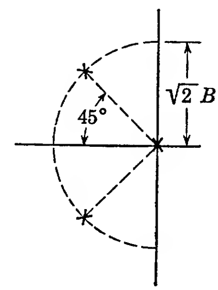


Fig. 14.9. Desired open-loop poles.

Figure 14.10b is characteristic of a high-pass filter if $(2)^{1/2}B$ is larger than the mean zero distance. In other words, it is a sophisticated kind of lead network. As $(2)^{1/2}B$ is made larger, the bandwidth with feedback B will be increased and the servo will be faster, at least up to the point where saturation begins to be troublesome.

Let us now describe the maximally flat servo in more detail. It will have the circuit of Fig. 14.5 and will be ruled by eqs. 14.9, 14.10, and

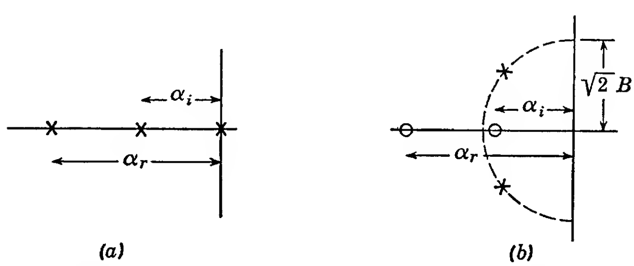


Fig. 14.10. Structure of the compensating function.

14.11, with the function F given by

$$F = \frac{2B^2(T_i p + 1)(T_r p + 1)}{p^2 + 2Bp + 2B^2} \qquad (14.26)$$

which is unity at zero frequency. When eq. 14.26 is substituted in eq.

14.11, the closed-loop function becomes

$$\frac{\theta_0}{\theta_i} = \frac{2B^2k_v}{p(p^2 + 2Bp + 2B^2) + 2B^2k_v} \tag{14.27}$$

which is maximally flat as desired if

$$k_v = kK_v/m = B/2 \tag{14.28}$$

which is the required loop gain of the system. The larger the closed-loop bandwidth B is made, the larger the gain can be made and the faster the servo will be, as long as saturation does not become a problem. In fact, the required loop gain is directly proportional to the closed-loop bandwidth.

The compensating function of Fig. 14.10*b* has poles with imaginary parts. Therefore, if the transfer function of the compensating network is to be obtained with bilateral or isolated networks, inductors must be employed. However, the use of inductors at the very low frequencies associated with servomechanisms is often impractical, which generally implies that the compensating function must be realized with a feedback amplifier. In the language of servomechanisms, the system would then be said to have a "minor loop," that is, a feedback network in cascade with the open-loop system. However, there is nothing mysterious about a minor loop, as it is but a direct consequence of obtaining an open-loop transfer function having complex poles without inductance in the compensating network.

Let us now see how these same techniques apply to the design of a fourth-order servo, that is, a servo having an open-loop transfer function with three poles in addition to the one at the origin. We shall assume the open-loop transfer function to be

$$\frac{\theta_0}{\epsilon} = \frac{k_v}{p(T_ip + 1)(T_rp + 1)(T_xp + 1)} \tag{14.29}$$

If T_x is small compared to T_i and T_r, then it may be possible to simply ignore T_xp compared to unity and design the system as has already been discussed. (Alternately, a lead network could be added to make T_x smaller.) We shall assume here that T_x cannot be ignored.

We want the closed-loop function to be maximally flat and have a bandwidth B. Therefore, we desire

$$\frac{\theta_0}{\theta_i} = \frac{B^4}{p^4 + 2.62Bp^2 + 3.42B^2p^2 + 2.62B^3p + B^4} \tag{14.30}$$

which requires the open-loop transfer function to be

$$\frac{\theta_0}{\epsilon} = \frac{2.62B^3k_v}{p(p^3 + 2.62Bp^2 + 3.42B^2p + 2.62B^3)} \tag{14.31}$$

which can be factored with numerical methods as

$$\frac{\theta_0}{\epsilon} = \frac{2.62B^3k_v}{p(p + 1.5B)(p^2 + 1.12Bp + 1.74B^2)} \tag{14.32}$$

The required value of k_v is seen to be

$$k_v = \frac{B^4}{2.62B^3} = 0.382B \tag{14.33}$$

which, for the same bandwidth, is somewhat less than that required for the simpler three-pole system. This, after all, was to be expected. The designer is always fighting phase shift in a feedback system and a four-pole open-loop transfer function is more of a problem in this regard than a three-pole transfer function. However, if T_x is small, a lag network will be required (that is, one pole will be brought closer to the origin) which permits more lead to be employed before getting into saturation difficulties. This makes the four-pole system almost as good as the three-pole system.

The required compensating network function F for this four-pole example must be

$$F = \frac{2.62B^3(T_ip + 1)(T_rp + 1)(T_xp + 1)}{(p + 1.5B)(p^2 + 1.12Bp + 1.74B^2)} \tag{14.34}$$

which has p-z as shown in Fig. 14.11 (where $\alpha_x = 1/T_x$). In no case is it advisable to use a compensating function having more poles than zeros as this but emphasizes the problem of lagging phase shift.

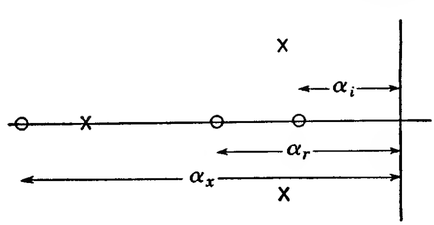


Fig. 14.11. Compensating function for a fourth-order system.

14.6 Compensating networks

There are innumerable feedback amplifiers that furnish p-z as in Fig. 14.10*b* without the use of inductance. The twin-T feedback amplifier, for example, was found to furnish two real zeros and one pair of complex-conjugate poles. The bridged-T feedback amplifier will also accomplish the desired result. As a specific example, consider the circuit of Fig. 14.12*a* in which assumptions similar to those made in connection with the study of the twin-T feedback amplifier of Chap. 11 are made

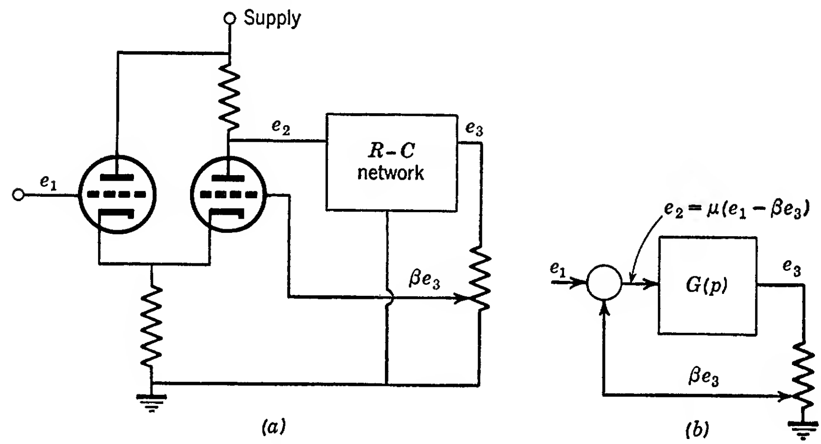


Fig. 14.12. A feedback amplifier for use in compensation.

so that $e_2 = \mu(e_1 - \beta e_3)$. The equivalent circuit of Fig. 14.12*a* is shown in Fig. 14.12*b*. If we designate the ratio e_3/e_2 by $G(p)$, then

$$\frac{e_2}{e_1} = \frac{\mu}{1 + \mu\beta G(p)}$$

$$\frac{e_3}{e_1} = \frac{\mu G(p)}{1 + \mu\beta G(p)} \tag{14.35}$$

We shall assume that one of the functions of eqs. 14.35 is to have the form of eq. 14.26, which is the compensating function $F(p)$ for the third-order servo. (Of course, the functions of eqs. 14.35 can be multiplied by any constant by means of a gain control.) It is therefore evident that the function $G(p)$ of eqs. 14.35 must have two and only two poles which are restricted to the negative real axis and which must be simple because $G(p)$ is to be an *R-C* network. If we take e_2 to be the output from the compensating network, then the poles of $G(p)$ must be at α_i

and α_r, whereas if we take e_3 to be the output, the zeros of $G(p)$ must be at α_i and α_r. If $G(p)$ is a function with poles at α_i and α_r but with no zeros (a two-section R-C low-pass filter), then the output must be e_2. However, if $G(p)$ does not have zeros, we cannot separately specify the bandwidth of the closed-loop system but must accept what the adjustment of the circuit of Fig. 14.12 calls for (although α_i and α_r can be made anything within reason by also using simple lead and/or lag networks in cascade with the open-loop system). Because this case is the

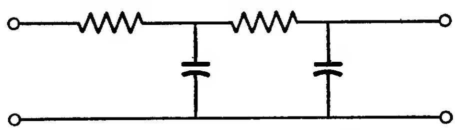


Fig. 14.13. The two-pole low-pass filter.

simplest, we shall study it first. Assume that $G(p)$ is given by the network of Fig. 14.13, which has a transfer function

$$\frac{e_3}{e_2} = G(p) = \frac{1}{(T_ip + 1)(T_rp + 1)} \tag{14.36}$$

The transfer function of the compensating feedback amplifier can then be found as

$$\frac{e_2}{e_1} = \frac{\mu(T_ip + 1)(T_rp + 1)}{(T_ip + 1)(T_rp + 1) + \mu\beta} \tag{14.37}$$

In order that the denominator of eq. 14.37 have poles at $-B \pm jB$, it is necessary that

$$\mu\beta = \frac{\alpha_r^2 + \alpha_i^2}{2\alpha_i\alpha_r}$$
$$B = \frac{\alpha_r + \alpha_i}{2} \tag{14.38}$$

A more versatile compensating network results by allowing the function $G(p)$ to have both p-z. Then, assuming e_2 is the output as in the previous example, $G(p)$ becomes

$$G(p) = \frac{ap^2 + bp + 1}{(T_ip + 1)(T_rp + 1)} \tag{14.39}$$

and the transfer function of the feedback amplifier is

$$\frac{e_2}{e_1} = \frac{\mu(T_ip + 1)(T_rp + 1)}{(T_ip + 1)(T_rp + 1) + \mu\beta(ap^2 + bp + 1)} \tag{14.40}$$

where the previous example was the special case of this more general expression for $a = b = 0$.

The relationships between the parameters of the denominator in order that eq. 14.40 have poles at $-B \pm jB$ are

$$\begin{aligned} \frac{T_i + T_r + \mu\beta b}{T_iT_r + \mu\beta a} &= 2B \\ \frac{1 + \mu\beta}{T_iT_r + \mu\beta a} &= 2B^2 \end{aligned} \tag{14.41}$$

For any given problem, T_i and T_r are known. Also, the bandwidth B is usually specified. This leaves a, b, and $\mu\beta$ as parameters to be determined. There are two equations but three parameters; therefore, either a, b, or $\mu\beta$ can be chosen. However, if $\mu\beta$ is arbitrarily chosen, the required function $G(p)$ may be nonminimum phase. Although a suitable network can be designed in this event, it is simpler to require that $G(p)$ be minimum phase. Then, the coefficients a and b will not be negative, which is assured if

$$\begin{aligned} 1 + \mu\beta &\geq 2B^2T_iT_r \\ 1 + \mu\beta &\geq B(T_i + T_r) \end{aligned} \tag{14.42}$$

As a specific example, assume $T_i = 1$, $T_r = 0.2$, and that the desired closed-loop bandwidth is $B = 10$ radians per second. Then, $\mu\beta$ must

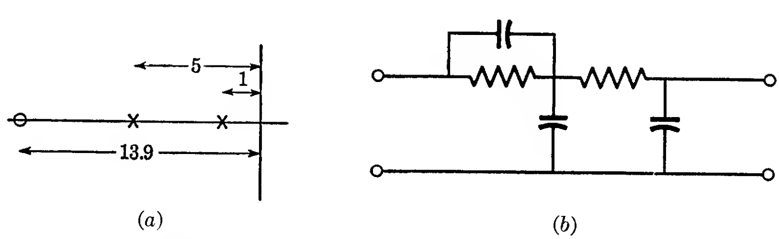


Fig. 14.14. The function $G(p)$ for the example.

not be less than 39. Let us take it to be equal to 39. Then, $a = 0$, $b = 0.0718$, and the resulting p-z of $G(p)$ are as shown in Fig. 14.14a. A network suitable for obtaining this $G(p)$ is shown in Fig. 14.14b.

14.7 Tachometer feedback

Tachometer feedback provides another means for putting the poles of the open-loop transfer function where they are desired. A tachometer is a speed measuring device fixed to the shaft of the servo motor. The output from the tachometer is a voltage proportional to the speed of the motor $p\theta_m$. If the tachometer is fed back around only the servo motor and is not injected into the major feedback loop, the closed-loop

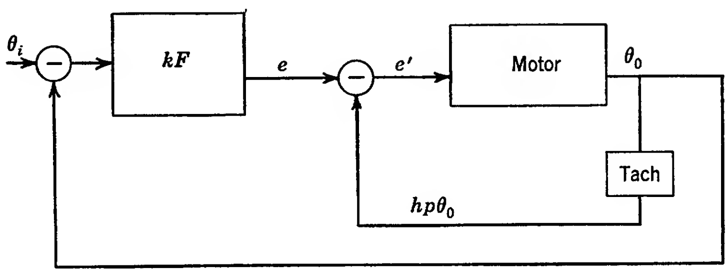


Fig. 14.15. A servo with tachometer feedback around the motor.

system is as shown in Fig. 14.15. Let us analyze a typical motor with tachometer feedback. If the transfer function of the motor alone is

$$\frac{\theta_0}{e'} = \frac{k_v}{p(T_ip + 1)(T_rp + 1)} \tag{14.43}$$

then the transfer function of the motor with tachometer feedback is

$$\frac{\theta_0}{e} = \frac{k_v}{p[(T_ip + 1)(T_rp + 1) + hk_v]} \tag{14.44}$$

from which it can be seen that a pole still exists at the origin, thus permitting the motor with tachometer feedback to be used within the feedback loop without violating the basic requirement of a servomechanism. Equation 14.44 may be written

$$\frac{\theta_0}{e} = \frac{k_v\alpha_i\alpha_r}{p[p^2 + p(\alpha_r + \alpha_i) + \alpha_r\alpha_i(1 + hk_v)]} \tag{14.45}$$

which can be made to have complex poles. In particular, if we wish it

to have poles at $-B \pm jB$, then

$$B = \frac{\alpha_i + \alpha_r}{2}$$
$$hk_v = \frac{\alpha_i^2 + \alpha_r^2}{2\alpha_r\alpha_i} \qquad (14.46)$$

Tachometer feedback around the servo motor is thus an alternative to a compensating network. However, whatever bandwidth results must be accepted.

Tachometer feedback may be placed in parallel with the feedback of the major feedback path rather than just around the motor in order to accomplish much the same thing. However, the closed-loop bandwidth can then be made anything within reason by introducing simple lead and lag networks in cascade with the open-loop system.

14.8 The effect of a load

Several different types of loads can be placed on the output motor of a servomechanism. The easiest type with which to deal is the ponderous load rigidly attached to the shaft of the servo motor, thereby adding additional moment of inertia and viscous friction. The behavior of a servo with this type of load is easily calculated by including the inertia and damping of the load as part of the armature inertia and damping. The system is analyzed accordingly, that is, as one having a larger moment of inertia and a larger damping than that given by the armature of the motor alone. Of course, the effect of a load must be translated to the armature of the motor if gearing is employed.

The more difficult type of load with which to deal is a varying torque load applied to the output shaft of the servo. This kind of load is essentially outside the servo feedback loop. To make matters more complex, the torque may be applied through elastic and massive members. We shall first consider the torque to be independent of the servo output angle and proceed to formulate an equivalent circuit.

Equation 14.6, suitably modified to account for a gear ratio m and an amplifier in cascade with the system having a transfer function kF, is

$$\theta_0 = \frac{(kFK_v/m)\epsilon - (R_aK_v/mk_t)(T_ap + 1)T_L}{p(T_ip + 1)(T_rp + 1)} \qquad (14.47)$$

where $T_a = L_a/R_a$, which may approximate T_r if the motor back voltage (as a generator) is small. The block diagram of the system described by this equation is shown in Fig. 14.16, which can be drawn directly

from an inspection of eq. 14.47. The system described by this diagram is quite similar to that of a feedback amplifier with an internal disturbance.

The closed-loop behavior of the system of Fig. 14.16 can be determined as

$$\theta_0 = \frac{k_vF}{p(T_ip + 1)(T_rp + 1) + k_vF}\theta_i - \frac{(R_ak_v/kk_t)(T_ap + 1)}{p(T_ip + 1)(T_rp + 1) + k_vF}T_L \tag{14.48}$$

where $k_v = kK_v/m$ is the effective velocity constant. It is assumed that

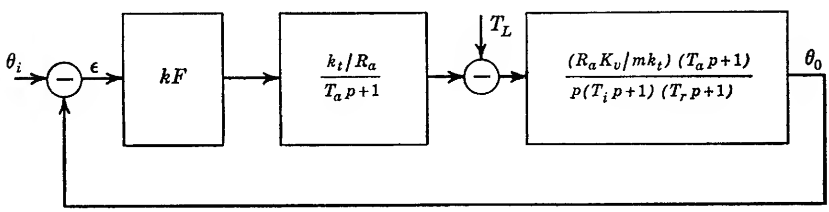


Fig. 14.16. Block diagram of a servo with a load.

F is unity at $\omega = 0$ as before. For $T_L = 0$, eq. 14.48 reduces to equations previously described.

Let us first study the behavior of the servo to independent load variations. The behavior to combined input and load variations can in this event be obtained by superposition. The first thing we note from eq. 14.48 is that increasing k, the open-loop gain *prior to* the point in the equivalent circuit where the load torque is introduced, reduces the severity of load torque effects. In other words, a high gain results in a large correcting torque for a small error. However, no matter how high k is (barring infinity), some error always results from a finite constant load torque.

With many servos, it is more important to minimize variations in θ_0 with load torque than it is to give faithful reproduction to input changes θ_i. It would then be expected that the design of the system should be based on a different criterion. First and most important is a large value of kk_t. In addition, the bandwidth of the system must be large enough to insure that corrections for changing load torques will be made without too great a delay. A large open-loop gain can be achieved by making the closed-loop bandwidth large. Although this may lead to saturation for large changes in the input θ_i, saturation may not at the same time be a serious factor for changes in T_L.

The compensation of servo open-loop functions for minimizing variations in load torque is more difficult than that for faithful reproduction of input signals. One thing that does help considerably is tachometer feedback around the entire system. Then the error becomes $\theta_i - \theta_0 - hp\theta_0$ instead of simply $\theta_i - \theta_0$. With tachometer feedback, eq. 14.47 leads to a closed-loop function (with $\theta_i = 0$ for simplicity in studying load effects) of

$$\theta_0 = \frac{-(T_ap + 1)(R_ak_v/kk_t)}{p(T_ip + 1)(T_rp + 1) + p(hk_vF) + k_vF} T_L \qquad (14.49)$$

A locus diagram of the denominator of eq. 14.49 shows how the system becomes more stable as the tachometer feedback factor h is increased.

So far, we have assumed a load torque T_L independent of θ_0 and θ_i. However, this is often not the case. As an example, consider the servo on a ship used to turn the rudder. The load torque is zero when the rudder is in the neutral position but increases as the rudder angle increases. If T_L is linearly dependent on θ_0, we can write $T_L = g\theta_0$, where g is a constant of proportionality. Equation 14.48 for this case becomes

$$\frac{\theta_0}{\theta_i} = \frac{k_vF}{p(T_ip + 1)(T_rp + 1) + k_vF + (gR_ak_v/kk_t)(T_ap + 1)} \qquad (14.50)$$

which can be made flat by means of compensating networks but which no longer has a zero error except in the trivial case when $\theta_0 = 0$.

The most difficult type of load torque to analyze occurs when a frequency-sensitive coupling exists between the servo and its load, for ex-

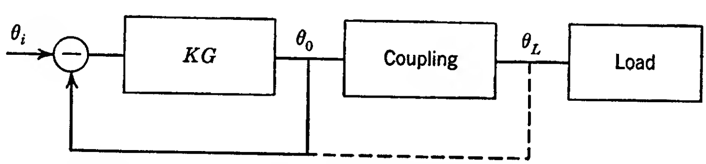


Fig. 14.17. Servo with coupling and load outside the loop.

ample, when a somewhat flexible shaft is used with a massive load. The circuit of such a system is shown in Fig. 14.17. The angle θ_L may then not be in phase with θ_0. The over-all transfer function is no longer completely controlled by feedback. What is perhaps more serious is that there will be a frequency-sensitive relationship between the torque on the servo motor and the angle θ_0, analogous to a feedback amplifier with a reactive load. This will not only affect the shape of the closed-loop function but might actually cause an otherwise stable system to

become unstable. The best procedure in such cases (although it may not often be possible) is to put the output sensing device at the *ultimate* load on the system as shown dashed in Fig. 14.17. Although this will complicate the open-loop function by introducing additional frequency-sensitive mechanical elements (mass and spring constants are analogous to inductors and capacitors), at least there is a more predictive control over the system.

14.9 The integrating servomechanism

It was found that the only way to have an output precisely equal to an input at very low frequencies and without load was to have a feedback system with an open-loop characteristic which shows a pole precisely at the origin of the p plane. If one pole at the origin gives such great benefits, perhaps two can give even more. Suppose an ideal integrator is placed in cascade with the open-loop system. Then, if any constant error exists because of a constant load torque, the error will be continually integrated until a correcting torque large enough to overcome the error is finally applied to the load by the servo motor. In other words, an integrating servo will not have a steady error when a constant load torque exists.

Let us study the simple integrating servo by assuming the function F of eq. 14.48 is given by F'/p. Of course, the meaning of the velocity constant k_v must then be modified; it affords a definite interpretation only if F is finite at $\omega = 0$. Equation 14.48 becomes

$$\theta_0 = \frac{k_vF'}{p^2(T_ip+1)(T_rp+1)+k_vF'}\theta_i - \frac{p(R_ak_v/kk_t)(T_ap+1)}{p^2(T_ip+1)(T_rp+1)+k_vF'}T_L \quad (14.51)$$

The factor p in the numerator of the second term of eq. 14.51 shows the low-frequency torque error to be zero, whereas the constant term

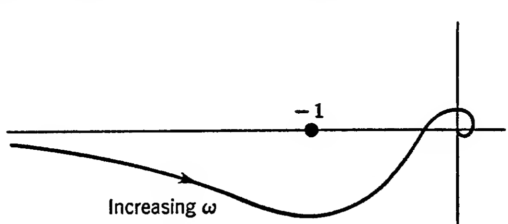


Fig. 14.18. Open-loop locus of an integrating servo.

k_vF' in both numerator and denominator of the first term of eq. 14.51 shows the no-load error to be zero, just as in the simple servo. Note, however, that if F' is a constant there is no term proportional to p in the denominator and hence the system will not be stable. In order to be stable, the open-loop locus must appear somewhat like that shown in Fig. 14.18. The shape of this locus requires that the open-loop phase shift be less than 180 degrees at medium frequencies. This in turn requires lead compensation, which can be obtained by making the function F have the form

$$F = \frac{Tp + 1}{p} \tag{14.52}$$

rather than simply $1/p$ (which is equivalent to having the function F' equal to $Tp + 1$). The existence of the factor $Tp + 1$ furnishes the missing power of p in the denominator of eq. 14.51.

The integrating servo can also be stabilized with F' a constant with tachometer feedback. The best procedure is to employ both lead compensation and tachometer feedback.

Of course, the integrator in cascade with the system can never be precisely realized with purely electric circuits. However, a precision analog computer integrating amplifier can be made to have a pole near enough to the origin to satisfy practical requirements.

14.10 Cluster functions in servo design

With feedback amplifiers, it was found that zeros could be added to to the open-loop transfer function of a system having only poles with beneficial results. The closed-loop behavior remains essentially that of an all-pole function, but the degraded bandwidth can be increased considerably as compared to that given by a simple one-pole bandwidth degrading function.

A similar situation is found in the servo with real feedback. However, the one-pole degrading function cannot be avoided; it is the single pole at the origin which defines a servomechanism. In feedback amplifiers, the one-pole degrading function can be replaced with a cluster of p-z with the number of poles exceeding the number of zeros by one. With a servo, a cluster of p-z can be added with the number of added poles equal to the number of added zeros. In neither case are the high-frequency poles changed. The addition of the cluster in feedback amplifiers permits the open-loop degraded bandwidth to be increased. In the servo, the addition of a cluster can increase or decrease the velocity constant without appreciably affecting the closed-loop gain characteristics.

Let a servo with an optimized all-pole closed-loop transfer function have the corresponding optimized open-loop transfer function

$$\frac{\theta_0}{\epsilon} = \frac{b_0 k_v}{p(p^n + b_{n-1}p^{n-1} + \cdots + b_1 p + b_0)} \tag{14.53}$$

where the velocity constant is k_v.

In cascade with this tailored open-loop function, let us place a suitable cluster function, which must asymptotically approach unity at high frequencies, and which must have the same number of p-z. It must therefore have the general form

$$\frac{p^m + a_{m-1}p^{m-1} + \cdots + a_1 p + a_0}{p^m + d_{m-1}p^{m-1} + \cdots + d_1 p + d_0} \tag{14.54}$$

With the function of eq. 14.54 in cascade with the open-loop characteristic of eq. 14.53, the velocity constant becomes

$$k_v' = \frac{k_v a_0}{d_0} \tag{14.55}$$

If the mean zero distance of the cluster $(a_0)^{1/m}$ is less than the mean pole distance $(d_0)^{1/m}$, the velocity constant is reduced and the servo is said to have "derivative control." Although a reduced velocity constant is not always desirable, derivative control does have the advantage of reducing the closed-loop phase lag.

If the mean zero distance of the cluster is greater than the mean pole distance, the effective velocity constant is increased and the servo is said to have "integral control." In fact, if $d_0 = 0$ in eq. 14.54, one of the poles of the cluster lies at the origin, the velocity constant goes to infinity, and an integrating servo results. In other words, the integrating servo is merely a special case of integral control.

With derivative control, the Nyquist plot at medium frequencies will tend to phase angles less than 90 degrees. With integral control and at medium frequencies, the phase shift is larger and the Nyquist plot will tend more to phase shifts approaching 180 degrees; in the limit of the integrating servo, the Nyquist plot of Fig. 14.18 results.

The number of p-z in the cluster function is dependent on the economics of the situation, that is, on how complex the designer is willing to make the compensating networks. Clearly, if a large number of p-z is used, the closed-loop characteristics of the servo can be made better. The simplest possible cluster function is obviously $(p + a_0)/(p + d_0)$, which is a simple lead or lag network function.

Problems

1. A simple servo with unity feedback has an open-loop transfer function $k_v/[p(p + 1)(0.1p + 1)]$. Determine the critical k_v. Plot the magnitude of θ_0/θ_i for k_v equal to $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{2}{3}$ the critical value.

2. A lag network $(0.1p + 1)/(0.4p + 1)$ is used with the simple servo of Prob. 1. Calculate the critical k_v and plot the magnitude of θ_0/θ_i for k_v equal to $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{2}{3}$ the critical value.

3. A lead network $(p + 1)/(0.5p + 1)$ is used with the simple servo of Prob. 1. Calculate the critical k_v and plot the magnitude of θ_0/θ_i for k_v equal to $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{2}{3}$ the critical value.

4. The lag network of Prob. 2 and the lead network of Prob. 3 are both used with the simple servo of Prob. 1. Calculate the critical k_v and plot the magnitude of θ_0/θ_i for k_v equal to $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{2}{3}$ the critical value.

5. A simple servo with unity feedback has an open-loop transfer function $k_v/[p(p + 1)(0.2p + 1)(0.05p + 1)]$. Calculate the critical k_v and plot the magnitude of θ_0/θ_i for k_v equal to $\frac{1}{3}$ and $\frac{2}{3}$ the critical value.

6. For the servo of Prob. 1, determine the numerical positions of the p-z of a compensating function such that the closed-loop transfer function will be maximally flat with a bandwidth of (*a*) 1 radian, (*b*) 2 radians, and (*c*) 3 radians. In each case, determine the critical gain and the required gain. Express the required gain as a percentage of the critical gain.

7. Plot the transfer function magnitudes versus frequency for the compensating networks of Prob. 6.

8. Repeat Prob. 6 using the Q-D function rather than the maximally flat function for the closed-loop behavior.

9. A special-purpose servo is to have a closed-loop transfer function $1/(0.5p + 1)^3$. Determine the p-z of the compensating function to work with the servo of Prob. 1. Determine the critical and required values of k_v.

10. Determine the p-z locations of a compensating network for the servo of Prob. 5 such that the closed-loop transfer function will be maximally flat and have a bandwidth of 2 radians.

11. The feedback amplifier of Fig. 14.12 uses a two-pole no-zero function $G(p)$ obtained with the network of Fig. 14.13. Determine the element values of the network such that, when used with the servo of Prob. 1, the closed-loop transfer function will be maximally flat. Determine the closed-loop bandwidth and the critical and required values of k_v.

12. A lead network $(p + 1)/(0.3p + 1)$ is used with the servo of Prob. 1. In addition, the feedback amplifier of Fig. 14.12 with the $G(p)$ of Fig. 14.13 is used. Repeat Prob. 11 for this case.

13. Design the lead network of Prob. 12 using two resistors and one capacitor to be driven from a cathode follower (that is, a voltage source).

14. Determine the element values of the circuit of Fig. 14.14*b* to give the p-z of Fig. 14.14*a*. Assume the network is driven from a voltage source.

15. A servo with its amplifier has an open-loop transfer function $k_v/[p(2p + 1)(0.2p + 1)]$. The over-all feedback is $\theta_0 + hp\theta_0$ obtained by means of a tachometer. Obtain the values of k_v and h so that the closed-loop transfer function will be maximally flat.

16. A lag network $(0.2p + 1)/(0.4p + 1)$ is placed in cascade with the open-loop system of Prob. 15. With this lag network, repeat Prob. 15.

17. A lead network $(2p + 1)/(p + 1)$ is placed in cascade with the open-loop system of Prob. 15. With this lead network, repeat Prob. 15.

18. Repeat Prob. 15 using a Q-D three-pole closed-loop transfer function and compare values.

19. A servo is designed to regulate against varying load torques with no interest in inputs θ_i. Tachometer feedback is used to give behavior as described by eq. 14.49. If $T_i = 1$, $T_r = 0.2$, $T_a = 0.3$, and $F = 1$, what is k_v so that the function θ_0/T_L will be fairly flat, and what is the bandwidth?

20. Suppose the function F in Prob. 19 and eq. 14.49 is a lead network $(p + 1)/(0.5p + 1)$. What value of k_v is required so that the function θ_0/T_L will be approximately flat and what is the approximate bandwidth?

21. A simple servo with three poles including the one at the origin has $k_v = 5$ and a maximally flat three-pole closed-loop transfer function with a (normalized) bandwidth of unity. Determine and plot the open-loop pole locations.

22. A cluster function $(p + 0.1)/(p + 0.05)$ is added to the open-loop system of Prob. 21. Determine the new velocity constant. Plot the resulting closed-loop gain function and compare with that of Prob. 21. It may be necessary to reduce the new value of k_v slightly in order to avoid a high-frequency peak. If so, what should this adjusted value be?

23. Repeat Prob. 22 for the cluster functions:

a. $(p + 0.2)/(p + 0.1)$.
b. $(p + 0.2)/(p + 0.05)$.
c. $(p + 0.1)/(p + 0.2)$.

Compare results to those of Probs. 21 and 22.

24. Repeat Prob. 22 for the cluster function

$$\frac{(p + 0.1)(p + 0.2)}{(p + 0.05)(p + 0.15)}$$

25. Add the cluster function $(p + 0.1)/p$ to the servo of Prob. 21. Determine the velocity constant and plot the closed-loop gain function.

References

The list of references given here is by no means intended to be complete; rather, it is merely representative of the background to this book. The source references to early and fundamental work by such men as Cauer, Foster, Campbell, and others have been included, although their contributions are adequately described in more recent texts in perhaps more readable form.

The presentation and approach used by the author in matters relevant to p-z represent a departure from standard methods. Consequently, in most cases the treatment in other books will look different from that here. However, if the reader looks between the lines of the references, similarities will be found. Perhaps the greatest difference between the author's treatment and that of others is the strict adherence to the phasor interpretation here; almost all the references that mention p-z develop the concepts from potential theory, contour integration, and analogs, or as a consequence of the Laplace transformation.

Articles in the current journal literature tend to be more similar to the methods employed by the author. However, only a few of these articles have been referenced because they are, for the most part, either not of general importance or are not germane to this book.

Chapter 1. Most books devoted to network analysis and designed for a second course in a-c circuit analysis include much of the material in Chap. 1. There are a great many of these texts, three of the more important being

1. E. A. Guillemin, *Communication Networks*, Vol. 2, John Wiley & Sons, New York, 1935.

2. E. A. Guillemin, *Introductory Circuit Theory*, John Wiley & Sons, New York, 1953.

3. W. R. Lepage and S. Seely, *General Network Analysis*, McGraw-Hill Book Co., New York, 1952.

The principal departure of the present book from refs. 1, 2, and 3 is the adoption of the superscript notation to avoid confusion between admittances and impedances and other quantities arising from nodal

and loop analyses. Also, the author has adopted the convention of showing the negative sign in the off-diagonal terms. Reference 2 is thorough in presenting graph theory as applied to electric networks. As contrasted to this book, the references cited do not lay as much emphasis on the manipulation of ideal transformers and determinants. A thorough review of determinants will be found in

4. E. A. Guillemin, *The Mathematics of Circuit Analysis*, John Wiley & Sons, New York, 1949.

Chapter 2. The treatment of functions in Chap. 2 is for the most part not similar to that in the literature listed here. General reference can therefore be made to books on college algebra and on elementary differential equations. A discussion of partial- and continued-fraction expansions will be found in many places, for example, ref. 4. Also, some information will be found in mathematics texts such as

5. R. V. Churchill, *Introduction to Complex Variables and Applications*, McGraw-Hill Book Co., New York, 1949.

Also, see the very important and comprehensive (but not too easy reading) book

6. H. W. Bode, *Network Analysis and Feedback Amplifier Design*, D. Van Nostrand Co., New York, 1945.

Bode's book contains a large number of circuit theorems, many of which are of extreme importance. His book (a classic) contains much original material.

Chapter 3. The rules regarding the permissible locations of p-z will be found in refs. 2, 3, 4, and 6. In particular, ref. 6 is rather complete. The author assumes responsibility for the various classifications of network functions.

Suggested books on transient analysis are

7. M. F. Gardner and J. L. Barnes, *Transients in Linear Systems*, John Wiley & Sons, New York, 1942.
8. E. Weber, *Linear Transient Analysis*, Vol. 1, John Wiley & Sons, New York, 1954.
9. S. Goldman, *Transformation Calculus and Electrical Transients*, Prentice-Hall, New York, 1949.

Heaviside's original work has been reprinted.

10. O. Heaviside, *Electromagnetic Theory*, Dover Publications, New York, 1950.

Chapter 4. The locations of the p-z of specific kinds of networks will be found in several of the references listed so far, in particular, 1, 2, and 6. Several factorization methods are rather thoroughly described

in ref. 8. It is generally conceded that modern network synthesis had its real beginnings with the study of reactance functions by Foster and Cauer. These historical references are

11. R. M. Foster, "A Reactance Theorem," *Bell System Tech. J.*, Vol. 3, 1924, pp. 259–267.

12. W. Cauer, "Die Verwirklichung von Wechslestromwiderständed Vorgeschriebener Frequenzabhängigkeit," *Arch. Elektrotech.*, Vol. 17, 1927, pp. 355–388.

Also see

13. T. C. Fry, "The Use of Continued Fractions in the Design of Electrical Networks," *Bull. Amer. math. Soc.*, Vol. 2, No. 35, July–Aug. 1929, pp. 463–498.

Treatments of impedance and frequency normalization are not often found in the literature in reasonably compact or complete form. The same can be said of frequency transformations. Reference 1 has some data.

Chapter 5. Discussions of the various kinds of gain functions are found widely scattered in the literature and are often merely presented with the assumption that the reader is already acquainted with the material. Actually, there are about as many functions as there are serious students of modern network theory, primarily because of the diversified criteria that lead to a definition of a "desirable" network function. Reference 3 has an elementary treatment of maximally flat and Chebyshev functions (which were originally developed by mathematicians rather than network theorists). Reference 1 discusses several functions as applied to lattice filters. Scattered material can also be found in ref. 6. Also see the references listed for Chap. 6. The treatment of the general Q-D function originates with the author.

14. J. L. Stewart, "Quasi-Distortionless Filter Functions," *Trans. I.R.E.*, PGCT-2, Dec. 1953, pp. 39–54.

Chapter 6. Material on modern network synthesis is continually coming out in the literature and several advanced textbooks are currently being written. Of major importance is

15. D. F. Tuttle, Jr., *Network Synthesis*, John Wiley & Sons, New York, 1957.

Also of fundamental importance is the article

16. E. A. Guillemin, "A Summary of Modern Methods of Network Synthesis," *Advances in Electronics*, Vol. 3, Academic Press, New York, 1951.

Also see

17. "Proceedings of the Symposium on Modern Network Synthesis," Polytechnic Institute of Brooklyn, New York, 1952.

In addition to these, a continuing series of transaction publications by the Professional Group on Circuit Theory of the Institute of Radio Engineers is of major importance. The papers published in these transactions in many cases include those given at National and Western conventions of the Institute of Radio Engineers. Convention papers may be published separately as part of the "Convention Record." Some of these papers contain excellent and comprehensive bibliographies, for example

18. S. Winkler, "The Approximation Problem of Network Synthesis," *Trans. I.R.E.*, CT-1, No. 3, Sept. 1954, pp. 5–20.

The design of the general maximally flat filter function is the author's own.

19. J. L. Stewart, "Shunt Capacitance and Maximally Flat Filter Design," *I.R.E. Convention Record*, Part 2, 1955, pp. 59–63.

A detailed study of ladder filters and their design for maximally flat, Chebyshev, or linear-phase behavior will be found in

20. E. Green, *Amplitude-Frequency Characteristics of Ladder Networks*, Marconi's Wireless Telegraph Co., Marconi House, Chelmsford, Essex (England), 1954.

Of major importance to modern network synthesis is the (difficult to read) article

21. S. Darlington, "Synthesis of Reactance 4-Poles which Produce Prescribed Insertion Loss Characteristics," *J. Math. and Phys.*, Vol. 18, No. 4, 1939, pp. 257–353.

A fundamental treatment of optimum mismatch using Chebyshev functions is given in

22. R. M. Fano, "Theoretical Limitations on the Broad-Band Matching of Impedances," *J. Franklin Inst.*, Vol. 249, Jan./Feb. 1950, pp. 57–83, 139–154.

The limitations introduced by shunt capacitance were first subjected to a critical analysis by Bode (ref. 6).

Additional references of importance (now mainly historical) are

23. C. M. Gewertz, *Network Synthesis*, Williams and Wilkins, Baltimore, 1933.
24. W. Cauer, *Theorie der Linearen Wechselstromschaltungen*, Becker and Erler, Leipzig, 1941. (Reprinted by Edwards Brothers, Ann Arbor, Mich., 1948.)

Chapters 7 and 8. The theory of image-matched filters dates back many years and is treated in a large number of texts. One of the most complete is ref. 1. Design data and tables will be found in

25. F. E. Terman, *Radio Engineer's Handbook*, McGraw-Hill Book Co., New York, 1943.

A recent and fairly comprehensive text on the subject is

26. M. B. Reed, *Electric Network Synthesis*, Prentice-Hall, New York, 1955.

A very early text on image-matched filters is

27. T. E. Shea, *Transmission Networks and Wave Filters*, D. Van Nostrand Co., New York, 1929.

The invention of image-matching concepts is attributed to Campbell and Zobel.

28. G. A. Campbell, "Physical Theory of Electric Wave Filters," *Bell System Tech. J.*, Vol. 1, No. 1, 1922.
29. O. J. Zobel, "Theory and Design of Uniform and Composite Electric Wave Filters," *Bell System Tech. J.*, Vol. 2, No. 2, 1923.

Chapter 9. The study of the properties of tubes with the aid of p-z nomenclature is evidently new. However, most of Chap. 9 is but a slight extension of methods employed by almost every author of books on electron-tube circuits, for example, ref. 25 and

30. F. E. Terman, *Electronic and Radio Engineering*, fourth edition, McGraw-Hill Book Co., New York, 1955.

The behavior of tubes at very high frequencies is considered in

31. G. E. Valley and H. Wallman, *Vacuum Tube Amplifiers*, M.I.T. Radiation Lab. Series, Vol. 18, McGraw-Hill Book Co., New York, 1948.
32. K. R. Spangenberg, *Vacuum Tubes*, McGraw-Hill Book Co., New York, 1948.

Chapters 10 and 11. Reference 31 is a fairly large book devoted entirely to vacuum-tube amplifiers and is one of the best references on this subject. Useful material can also be found in refs. 3, 25, and 30. An elementary treatment somewhat similar to that in this book will be found in

33. T. L. Martin, Jr., *Electronic Circuits*, Prentice-Hall, New York, 1955.
34. M. E. Van Valkenberg, *Network Analysis*, Prentice-Hall, New York, 1955.

Credit for the discovery of the maximally flat amplifier, and the original basis for the design of isolated systems on a precision basis, should go to Butterworth.

35. S. Butterworth, "On the Theory of Filter Amplifiers," *Exp. Wireless*, Vol. 7, Oct. 1930, pp. 536–541.

The importance of Butterworth's work seems to have been overlooked for a long time, until rediscovered and extended in 1941.

36. V. D. Landon, "Cascade Amplifiers with Maximal Flatness," *RCA Rev.*, Vol. 5, Jan. 1941, pp. 347–362, and Vol. 5, Apr. 1941, pp. 481–497.

Chapter 12. The most complete available treatment of feedback amplifiers is ref. 6, in which the log-frequency design procedure (originated by Bode) is described in detail. In addition, almost every text on electron-tube circuits has some material on feedback, for example, refs. 25 and 30. However, none make free use of the phasor interpretation of p-z. The precision p-z method, including the concept of the cluster function, originates with the author.

Data concerning analog computer amplifiers can be found in

37. I. A. Greenwood, J. V. Holdam, and D. Macrae, *Electronic Instruments*, M.I.T. Radiation Lab. Series, Vol. 21, McGraw-Hill Book Co., New York, 1948.

38. G. A. Korn and T. M. Korn, *Electronic Analog Computers*, McGraw-Hill Book Co., New York, 1952.

Chapter 13. The author's treatment of oscillators is somewhat different in detail from that described in the references; in particular, it is more generalized. A comprehensive study of oscillators and an extensive bibliography on the subject will be found in

39. W. A. Edson, *Vacuum Tube Oscillators*, John Wiley & Sons, New York, 1953.

The behavior of oscillators at very high frequencies is analyzed in detail in

40. D. R. Hamilton, J. K. Knipp, and J. B. H. Kuper, *Klystrons and Microwave Triodes*, M.I.T. Radiation Lab. Series, Vol. 7, McGraw-Hill Book Co., New York, 1948.

Studies of electronic tuning will not be found elsewhere except as applied to specific devices. The parallel-network oscillator originates with the author.

41. J. L. Stewart, "Parallel-Network Oscillators," *Proc. I.R.E.*, Vol. 43, No. 5, May 1955, pp. 589–595.

Chapter 14. This chapter on servomechanisms is just as applicable to feedback amplifiers. The treatment is considerably different from that found in most of the references, although the approach is in some respects similar to

42. M. R. Aaron, "Synthesis of Feedback Control Systems by Means of Pole and Zero Locations of the Closed Loop Function," *Trans. AIEE*, Vol. 70, 1951, pp. 1439–1446.

43. J. G. Truxal, *Automatic Feedback Control System Synthesis*, McGraw-Hill Book Co., New York, 1955.

Most of the important work pertaining to servomechanisms is reported in the Transactions of the American Institute of Electrical Engineers.

Other general references to servomechanisms are not similar to Chap. 14, although they are many and varied. Perhaps of greatest value are

44. G. S. Brown and D. P. Campbell, *Principles of Servomechanisms*, John Wiley & Sons, New York, 1948.

45. H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design*, Vol. 2, John Wiley & Sons, New York, 1955.

46. H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms*, M.I.T. Radiation Lab. Series, Vol. 25, McGraw-Hill Book Co., 1947.

47. H. Lauer, R. Lesnick, and L. E. Matson, *Servomechanism Fundamentals*, McGraw-Hill Book Co., New York, 1947.

Index